Tennessee Law Review

Volume 59
Issue 1 *1991-1992*

Article 1

1992

Volume 59

TENNESSEE LAW REVIEW

VOLUME 59

1991-1992

TENNESSEE LAW REVIEW

VOLUME 59 | FALL 1991 | NUMBER 1

CONTENTS

TENNESSEE LAW REVIEW

Volume 59 Winter 1992 Number 2

CONTENTS

TENNESSEE LAW REVIEW

VOLUME 59 | SPRING 1992 | NUMBER 3

CONTENTS

SYMPOSIUM ISSUE: RECENT WORK IN FEMINIST LEGAL THOUGHT

RECENT WORK IN FEMINIST LEGAL THOUGHT

INTRODUCTION

MARILYN V. YARBROUGH*

When invitations were issued for participation in this symposium, scholars were told only that a special issue of the *Tennessee Law Review* was being planned, featuring the text of a speech by Professor Robin West. The typescript of the speech was not yet available, and the single paragraph that described the lecture indicated simply that Professor West had criticized the classically liberal vision of "ordered liberty." Her criticism was of courts' failure to take women's experiences and needs into account by relying on a concept that recognized a liberty interest only in the right to freedom from state interference with certain actions. In the speech she called for a reconstruction of a better and richer vision of liberty.

Indicating that the West manuscript would be available by the end of the year, the faculty advisors to the symposium issue asked that other symposium issue submissions arrive at the Law Review office within a month of that time. The limited time and sketchy description[1] may have discouraged some who received invitations, but not the authors of the works that follow. In what emerges as a remarkably cohesive group of articles, essays, speeches, and short stories, these authors have pushed, prodded, nudged, and shoved us toward the reconstruction demanded by Professor West.

In the lead article,[2] Professor West argues that not only does the modern liberal vision of "ordered liberty,"[3] fail to fully consider the

---

* Visiting Maier Chair of Law, West Virginia University; Professor of Law, University of Tennessee.

1. Of course, it is quite likely that those who received the invitation were already familiar with an earlier work of Professor West in which she discusses the differences between the approaches taken by Justice Scalia and Justice Brennan—Robin West, *The Ideal of Liberty: A Comment on* Michael H. v. Gerald D., 139 U. OF PENN. L. REV. 1373 (1991)—and so had some familiarity with her criticism of the concept of "ordered liberty." They were not aware, however, of the more specific outlines of her speech.

2. Robin West, *Reconstructing Liberty*, 59 TENN. L. REV. 441 (1992).

3. The term was first coined by Justice Cardozo in the 1930s and more fully explored in several Warren era decisions. *Id* at 442.

most fundamental of women's needs and experiences, it is not consistent with the history of the Fourteenth Amendment.[4] The construction to which she refers, freedom from state interference with certain personal decisions, considers liberty "the right to be left alone, not . . . right[s] to any particular way to be."[5]

In contrast to this "negative" construction, she asserts that the Fourteenth Amendment "must be understood as including [the] *positive* rights of autonomy, economic self-sufficiency, and political self-governance."[6] Referring to the abolitionist purposes of the Fourteenth Amendment and examining federal legislation passed just after the Amendment was ratified, she argues convincingly that the concern of the day was for the end to brutal and discriminatory *private* action exacerbated by state inaction.[7]

In addition, Professor West describes the current concept of ordered liberty as not only inconsistent with the origins of the Fourteenth Amendment, but "[i]t positively protects the sphere of privacy, negative liberty, and individual freedom within which women are most vulnerable and within which women are uniquely, individually, and definitively oppressed."[8]

Distinguishing her criticism from that of Justice Scalia and other conservatives who would adopt the "negative freedom" and "state action" prohibitions implicit in the liberal interpretation, but limit them to those rights allegedly historically and traditionally protected,[9] West uses two examples—constraints imposed by unequal domestic responsibilities and the threat of sexual violence—to fashion a convincing argument that a reconstruction of liberty that would meet the important needs of women can be accomplished without doing violence to the Constitution.

Responding to the invitation to address this issue, the other participants in this symposium have addressed subjects as diverse as the career choices made by a young woman lawyer[10] and euthanasia.[11] For perspective, we first present a delightful essay, *The Coming American Woman*,[12] written in late 19th century, predicting a "more

---

4. *Id.* at 465.
5. *Id.* at 447.
6. *Id.* at 466 (emphasis added).
7. *Id.* at 467 (emphasis added).
8. *Id.* at 461.
9. *Id.* at 444-45. To the extent that sodomy laws, prohibitions against homosexuality, abortion, and non-recreational drug use existed in the past, these would not be considered historically and traditionally protected.
10. Mary F. White, *Women in the Law*, 59 TENN. L. REV. 577 (1992).
11. Leslie Bender, *A Feminist Analysis of Physician-Assisted Dying and Voluntary Active Euthanasia*, 59 TENN. L. REV. 519 (1992).
12. Emily Sloan, *The Coming American Woman*, *in* 59 TENN. L. REV. 469 (1992).

equalized life" for women, reminiscent of the initial partnership of Adam and Eve, before, as Emily Sloan describes it, "Eve played hookey."[13] Professor Bari Burke, in her "afterword" to this essay, echoes the theme first sounded in Robin West's centerpiece and repeated throughout this symposium, women's search for selfhood, identity, and autonomy, as applied to Emily Sloan.

The next two articles address the necessity for autonomy in medical treatment. Professor Lisa Ikemoto, in her discussion of forced medical treatment of pregnant women,[14] sounds a theme that is found in yet another symposium article by Professor Frances Ansley,[15] that when we ignore issues of race and class and their intersection with issues of gender, we fail to fully understand the nature of patriarchy. Like Professor Ansley, she prods us to build "choice from coalition."[16]

In the second medical treatment article, Professor Leslie Bender offers a model for examining questions raised when competent patients request physicians' assistance in dying.[17] Like several of the other authors, she employs narrative to focus our attention on the ethical, moral, medical, and legal issues involved in this very difficult question, as significant for men as for women. In this context she explores the issues of autonomy raised in each of the other articles.

In an argument reminiscent of Professor West's concerns about our present conception of liberty as "negative" or "positive," Professor Bender decries the problem of "false dualism" that causes us to categorize physicians' actions as either passive or active and to legally ignore the former while condemning the latter. She argues for a construction that would consider patients' needs, concerns, and values just as Professor West and others argue for a general construction of liberty that would consider the needs of women.

Professor Teresa Godwin Phelps, a self-described "Catholic feminist legal scholar,"[18] writes about her search, and that of other Catholic women, for a method of breaking the silence to which they have traditionally resorted on the issue of abortion. Her interviews with ten Catholic professional women form the core of this article and confirm that "[a]bortion is a matter that is morally problematic, pastorally delicate, legislatively thorny, constitutionally insecure,

---

13. *Id.* at 470.

14. Lisa C. Ikemoto, *Furthering the Inquiry: Race, Class, and Culture in the Forced Medical Treatment of Pregnant Women*, 59 TENN. L. REV. 487 (1992).

15. Frances Lee Ansley, *A Civil Rights Agenda for the Year 2000: Confessions of an Identity Politician*, 59 TENN. L. REV. 593 (1992).

16. Ikemoto, *supra* note 14, at 488.

17. Bender, *supra* note 11.

18. Teresa Godwin Phelps, *The Sound of Silence Breaking: Catholic Women, Abortion, and the Law*, 59 TENN. L. REV. 547 (1992).

ecumenically divisive, medically normless, humanly anguishing, racially provocative, journalistically abused, personally biased, and widely performed."[19]

Mary F. White has contributed two provocative short stories, *Choice*,[20] and *Women in the Law*,[21] the first sounding remarkably as if the conversation had taken place with one of Professor Phelps' interviewees as it explores the complex considerations surrounding the decision of whether or not to abort a fetus, and the second, addressing the problems facing a young female lawyer as she considers the choices presented by personal relationships and professional expectations. The voices of these women as they wrestle with these complicated issues introduce a distinctly human dimension to the dominant theme, the achievement and exercise of autonomy.

In a critical broad look to the future, Professor Fran Ansley addresses the theme of "identity," exploring concepts appropriate for a civil rights agenda for the 21st Century.[22] In her article, she speaks, not of individual autonomy, but of the politics of race and gender, observing that participants in recent reform movements coalesce around group identity.[23] For all of the good that has come from these reform movements, they often spend precious time and resources on defining who is in and who is out.[24] Reminding us that we all fall into multiple categories, she urges us to "see beyond the lens of [our] own group identification,"[25] and when engaging in political activity based on our identity with that group, to consider the perspective and needs of those in the group who are least privileged and therefore less likely to achieve autonomy without our assistance.

Professor Stephanie Levin addresses a slightly different topic,[26] but returns to the theme of anti-subordination and self-determination that is so pervasive in these articles. She questions whether, as women are permitted to enter what she describes as "the previously all-male domains of violence," we will gain a new perspective on peace, citing our propensity to conform to, rather than transform, public institutions.[27] She too urges a reconstruction: a reconstruction of the way

---

19. *Id.* at 549 (quoting RICHARD A. MCCORMICK, S.J., HOW BRAVE A NEW WORLD?: DILEMMAS IN BIOETHICS 118-19 (1981)).

20. Mary F. White, *Choice*, 59 TENN. L. REV. 571 (1992).

21. Mary F. White, *Women in the Law*, 59 TENN. L. REV. 577 (1992).

22. Ansley, *supra* note 15.

23. *E.g.*, the "women's movement," the "gay and lesbian movement." Ansley, *supra* note 15, at 599.

24. *Id.* at 600.

25. *Id.* at 607.

26. Stephanie A. Levin, *Women, Peace, and Violence: A New Perspective*, 59 TENN. L. REV. 611 (1992).

27. She cites her and others' intention upon entering law school to make the

we think about public citizenship or patriotism that does not have as its defining principle, "the right to fight."

The final article in this symposium is a student comment on sex-specific fetal protection policies.[28] After examining fetal protection policies including legislation, lower court decisions, commentary, and debates among feminist legal theorists regarding "equal treatment" versus "special treatment," Jennifer Morton reiterates Robin West's suggestion that the policies of private actors that systematically infringe on women's liberty be subject to constitutional challenge under the Fourteenth Amendment: that "'affirmative protection' legislation," not unlike that already found in Title VII, be mandated when a working mother's *or father's* interests and needs in "conceiving, birthing, and rearing" children is threatened, thus protecting women's liberty to choose their destinies by choosing their careers.

This remarkable collection of women's work about women's liberty, wrestling with vexing, complicated issues, is much more than an initial step toward reconstructing present notions of liberty; it provides a blueprint for the reexamination of issues of self-determination and autonomy affecting women and men.

---

legal profession less "competitive," "pugnacious," and "brutal," conceding some adaptation of the profession to the presence of women, but acknowledging the tendency of many to conform. *Id.* at 615.

28. Jennifer Morton, Comment, *Pregnancy in the Workplace—Sex-Specific Fetal Protection Policies*—UAW v. Johnson Controls, Inc.—*A Victory for Women?*, 59 TENN. L. REV. 617 (1992).

TENNESSEE LAW REVIEW

VOLUME 59 | SUMMER 1992 | NUMBER 4

CONTENTS

SYMPOSIUM: ENVIRONMENTAL RIGHTS AND INTERNATIONAL PEACE

DENYING A RIGHT BY DISREGARDING DOCTRINE: HOW *ILLINOIS V. RODRIGUEZ* DEMEANS CONSENT, TRIVIALIZES FOURTH AMENDMENT REASONABLENESS, AND EXAGGERATES THE EXCUSABILITY OF POLICE ERROR

THOMAS Y. DAVIES*

---

* Associate Professor of Law, University of Tennessee College of Law. B.A., 1969, University of Delaware; J.D., 1975, Ph.D. (Political Science), 1980, Northwestern University. The author expresses his thanks to Professors Neil P. Cohen and Joseph R. Cook for reading and commenting on the manuscript for this article. Of course, all opinions and any errors that remain are the sole responsibility of the author.

> [W]e are in danger of forgetting that the Bill of Rights reflects experience with police excesses. It is not only under Nazi rule that police excesses are inimical to freedom. It is easy to make light of insistence on scrupulous regard for the safeguards of civil liberties when invoked on behalf of the unworthy. It is too easy. History bears testimony that by such disregard are the rights of liberty extinguished, heedlessly at first, then stealthily, and brazenly in the end.[1]

I. INTRODUCTION

There is a vision of order in recent United States Supreme Court decisions involving Fourth Amendment search and seizure issues. The Rehnquist Court has a pronounced tendency to uphold aggressive police conduct.[2] Two factors account for the predictability of the

---

1. Davis v. United States, 328 U.S. 582, 597 (1946) (Frankfurter, J., dissenting).

2. In the 1990 Term, the Court decided five search or seizure cases, all in favor of the government: California v. Hodari D., 111 S. Ct. 1547 (1991) (drugs thrown by the defendant a moment before he was tackled by a police officer who was chasing him without cause are admissible because no Fourth Amendment seizure had yet occurred at the precise moment in time when the drugs were thrown); County of Riverside v. McLaughlin, 111 S. Ct. 1661 (1991) (person subjected to warrantless arrest may be held up to 48 hours before being presented to a magistrate for assessment of probable cause for arrest); Florida v. Jimeno, 111 S. Ct. 1801 (1991) (defendant's consent to search of car for drugs constituted consent to search of contents of containers found in car); California v. Acevedo, 111 S. Ct. 1982 (1991) (a warrantless search of the contents of a paper bag that the police had probable cause to believe held marijuana was constitutional because the bag had been placed in the trunk of a car); Florida v. Bostick, 111 S. Ct. 2382 (1991) (consent given by seated bus passenger to a search of his luggage after police boarded bus and stood in aisle while requesting such consent does not require inference that consent was involuntary).

For similar decisions from previous terms, see Yale Kamisar, *Remembering the "Old World" of Criminal Procedure: A Reply to Professor Grano*, 23 U. MICH. J.L. REF. 537, 553 n.60 (1990).

For commentary that documents the tendency of the Rehnquist and Burger Courts to uphold aggressive police conduct, see Yale Kamisar, *The "Police Practice" Phases of the Criminal Process and the Three Phases of the Burger Court, in* THE BURGER YEARS: RIGHTS AND WRONGS IN THE SUPREME COURT, 1969-1986 (Herman Schwartz ed., 1987); Laurence A. Benner, *Diminishing Expectations of Privacy in the Rehnquist Court*, 22 J. MARSHALL L. REV. 825 (1989); John M. Burkoff, *The Court that Devoured the Fourth Amendment: The Triumph of an Inconsistent Exclusionary Doctrine*, 58 ORE. L. REV. 151 (1979); Yale Kamisar, Gates, *"Probable Cause," "Good Faith," and Beyond*, 69 IOWA L. REV. 551 (1984); Wayne R. LaFave, *The Forgotten Motto of Obsta Principiis in Fourth Amendment Jurisprudence*, 28 ARIZ. L. REV. 291 (1986); Wayne R. LaFave, *Fourth Amendment Vagaries (of Improbable Cause, Imperceptible Plain View, Notorious Privacy, and Balancing Askew)*, 74 J. CRIM. L. & CRIMINOLOGY 172 (1983); Tracy Maclin, *Constructing Fourth Amendment Principles from the Government Perspective: Whose Amendment Is It, Anyway?*, 25

direction of these decisions. The first is the membership of the Court. It is well understood that a principal determinant of Supreme Court decisions is the membership of the Court,[3] and the voting patterns of the Justices reveal that, for at least the last two decades, there has been a bloc of at least five Justices who have tended to favor aggressive police conduct in search cases.[4] The predictability of the Justices' voting positions is not surprising; an ideological predilection to favor expansive police power in such cases usually has been a political requirement for nomination to the Court during recent administrations.[5]

AM. CRIM. L. REV. 669 (1988); William J. Mertens, *The Fourth Amendment and the Control of Police Discretion*, 17 MICH. J.L. REF. 551 (1984); Stephen A. Saltzburg, *Another Victim of Illegal Narcotics: The Fourth Amendment As Illustrated by the Open Fields Doctrine*, 48 PITT. L. REV. 1 (1986); Silas J. Wasserstrom, *The Incredible Shrinking Fourth Amendment*, 21 AM. CRIM. L. REV. 257 (1984); Larry W. Yackle, *The Burger Court and the Fourth Amendment*, 26 U. KAN. L. REV. 335 (1978).

3. *See, e.g.*, DAVID W. ROHDE & HAROLD J. SPAETH, SUPREME COURT DECISION MAKING 157 (1976) (reporting that approximately 86% of the votes cast by Justices in cases decided between 1958 and 1973 could be predicted on the basis of models of the Justices' values). *See also* J. HARVIE WILKINSON, III, SERVING JUSTICE: A SUPREME COURT CLERK'S VIEW 146-47 (1974) ("Criminal rights . . . may . . . be the part of the Court's work most susceptible to swings of the pendulum after a change of personnel . . . [especially in view of] the 'law and order' criteria for the new appointees . . . .").

4. With the retirement of Justices Brennan and Marshall, the predictable majority for progovernment rulings in search cases has increased to at least six, possibly seven. Justice Stevens is the only sitting justice who has shown any consistent willingness to vote against the government in Fourth Amendment cases. Justice Blackmun has shown more willingness to vote against the government in Fourth Amendment cases than any of the other Nixon appointees, but he usually votes in favor of the government.

5. The Nixon administration, which nominated Chief Justice Burger and Justices Blackmun, Powell, and Rehnquist, along with two rejected judicial nominees, Judges Carswell and Haynsworth, screened potential nominees on the basis of their adherence to law and order ideology. There was a temporary respite from this process when President Ford named an ideological moderate, Justice Stevens, in the immediate aftermath of the Watergate scandals. No vacancy occured during the Carter administration. The Reagan administration, which nominated Justice Rehnquist to be Chief Justice, and which also nominated Justices O'Connor, Scalia and Kennedy, along with one rejected nominee, Judge Bork, engaged in even more vigorous ideological screening than did the Nixon Administration. Of the nine justices sitting during the 1989 Term, five were nominated by the Nixon or Reagan administrations.

It appears the Bush Administration, which has sucessfully nominated Justices Souter and Thomas to replace Justices Brennan and Marshall, is continuing to screen nominees according to much the same criteria employed in the Reagan administration. Justice Souter voted for the government in all five of the search and seizure cases decided in his first term on the Court. *See* cases cited *supra* note 2. Hence, six of the Justices sitting during the 1990 term were selected in part for their crime control ideological outlook; this number rose to seven in the 1991 term.

For a defense of the Court-packing practiced in the Nixon and Reagan administrations, see WILLIAM H. REHNQUIST, THE SUPREME COURT: HOW IT WAS, HOW IT IS 234-51 (1987).

The second factor that explains the progovernment direction of search decisions may not be as apparent. It may even be obscured by the widespread currency of the notion that criminal suspects often escape conviction because legal technicalities derail prosecutions. This notion might suggest the reason the majority Justices can vote so consistently for the government is because there is a large supply of cases in which lower court judges have released defendants accused of serious crimes because of arcane, or inane, legal technicalities. This is not the case, however. The reality is that due process standards are only a marginal factor in the criminal disposition process.[6]

The second reason why the majority Justices can vote so consistently for the government in search cases is because they too often disregard constitutional doctrine and principles. My complaint is not about the majority Justices simply making choices in the interstices and penumbras of legal doctrine. It is beyond argument that constitutional doctrine is not a closed or determinate system and that some degree of choice is inevitable. Given its membership, it is to be expected that the present Court will make interstitial choices that will favor the state.

My complaint is directed, rather, to ideological choices of an entirely different magnitude: the majority Justices decline to be bound

---

6. *See* Thomas Y. Davies, *A Hard Look at What We Know (and Still Need to Learn) About the "Costs" of the Exclusionary Rule: The NIJ Study and Other Studies of "Lost" Arrests*, 1983 AM. BAR F. RES. J. 611. I have estimated that only between 0.6% and 2.35% of all felony arrests are lost during the stages of the criminal dispostion process (prosecutorial screening, motions to suppress in courts, and appeals) as a result of illegal search problems or rulings. *Id.* at 621, 679-80. This estimate was discussed in both the majority and a dissenting opinion in United States v. Leon, 468 U.S. 897, 907 n.6 (majority opinion), 942, 950 n.11 (1984) (Brennan, J., dissenting). This estimate was also adopted in SPECIAL COMMITTEE ON CRIMINAL JUSTICE IN A FREE SOCIETY, CRIMINAL JUSTICE SECTION, AMERICAN BAR ASSOCIATION, CRIMINAL JUSTICE IN CRISIS 16-17 (1988). This committee report concludes "Constitutional restrictions, such as the exclusionary rule and Miranda, do not significantly handicap police and prosecutors in their efforts to arrest, prosecute, and obtain convictions of criminal defendants for most serious crimes." *Id.* at 5. It also reports "the prosecutors and police the Committee interviewed . . . do not believe that Fourth Amendment rights are a significant impediment to crime control. The exclusionary rule affects only a relatively small percentage of arrests and seizures." *Id.* at 8. *See also* Craig D. Uchida & Timothy S. Bynum, *Search Warrants, Motions to Suppress and "Lost Cases": The Effects of the Exclusionary Rule in Seven Jurisdictions*, 81 J. CRIM. L. & CRIMINOLOGY 1034, 1064 (1991) (reporting results consistent with Davies, *supra*).

Even the United States Department of Justice now agrees "[t]here is some consensus that the exclusionary rule costs the state only a small percentage of the total of all possible felony prosecutions . . . ." OFFICE OF LEGAL POLICY, U.S. DEPT. OF JUSTICE, REPORT TO THE ATTORNEY GENERAL ON THE SEARCH AND SEIZURE EXCLUSIONARY RULE, "TRUTH IN CRIMINAL JUSTICE" REPORT NO. 2 (1986), *reprinted in* 22 MICH. J.L. REF. 573, 609 (1989).

For discussions of some of the reasons why due process rules have only a minimal effect on the disposition of felony arrests, see sources cited *infra* note 9.

by, or even much influenced by, fundamental legal concepts and the basic principles and values expressed in the Fourth Amendment. Moreover, the majority Justices sometimes have been less than candid regarding their disregard of existing doctrine, neither acknowledging nor explaining their doctrinal revisions. Instead, they simply have asserted radical doctrinal claims while feigning continuity with earlier traditions.

The willingness of the majority Justices to disregard legal doctrine is important because the enforceability of a right depends on the coherence and the content of the doctrine that defines that right, as well as on the degree to which that doctrine produces clear boundaries in the form of rules. Consequently, the Fourth Amendment right to be secure against unreasonable government invasions of privacy is an enforceable reality only to the degree that the notion of an unreasonable search has operative doctrinal content.

In the past, the Court has understood Fourth Amendment reasonableness in a way that did have a substantial operative meaning: a government intrusion had to be justified in terms of probable cause and usually had to be authorized by a judicially issued warrant. The Rehnquist Court, however, ignores that meaning and tends to read the reasonableness requirement in such a loose and formless way that enforcement of the right announced in the Fourth Amendment is greatly diminished. The Rehnquist Court tends to treat Fourth Amendment reasonableness as a flexible, ad hoc, colloquialized notion.

It appears, in fact, that a majority of the Justices have embarked on a campaign to replace the basic idea that the Fourth Amendment sets out an enforceable right with the notion that the amendment merely authorizes the courts to regulate, as they think advisable, the most blatant instances of police misconduct.[7] This regulatory understanding of the Fourth Amendment, however, merely gives reviewing courts discretion to rein in government intrusions[8] if and when they choose to do so; it does not require reviewing courts to enforce a citizen's right. This difference is important because there are reasons to think that reviewing courts are generally reluctant to suppress

---

7. *See* Maclin, *supra* note 2 (arguing the Court has increasingly adopted a "police perspective analysis" when deciding Fourth Amendment issues); *cf.* Anthony G. Amsterdam, *Perspectives on the Fourth Amendment*, 58 MINN. L. REV. 349, 367 (1974) (discussing the replacement of a focus on the personal rights of individuals with "essentially a regulatory canon").

8. I use the term "intrusion" as a short-hand term for all of the varieties of police conduct that can implicate Fourth Amendment rights. The term is particularly useful in the context of Illinois v. Rodriguez, 110 S. Ct. 2793 (1990), because the police conduct at issue in that case consisted of an entry of a residence for the purpose of making an arrest of the resident for battery. The police conduct could be described as a "search," but it was not a search as that term is commonly used.

unconstitutionally seized evidence in any event.[9] Search standards that

---

9. There is reason to think reviewing courts are reluctant to suppress evidence, and will avoid suppression if search standards are elastic enough to permit them to do so. This is a sensible expectation because the only mechanism that has been developed to enforce the Fourth Amendment is the exclusionary rule. Motions to suppress under the exclusionary rule are made after incriminating evidence has been seized. There is tangible pressure on reviewing courts, at least if the crime is serious, to uphold the search or seizure. *See* LYNN M. MATHER, PLEA BARGAINING OR TRIAL: THE PROCESS OF CRIMINAL CASE DISPOSITION 73, 146 (1979) (trial judges routinely accept police testimony in disputes about circumstances of searches); JEROME H. SKOLNICK, JUSTICE WITHOUT TRIAL: LAW ENFORCEMENT IN A DEMOCRATIC SOCIETY 221 (2d. ed. 1975) ("The illegality of the search is likely to be tempered, even in the eyes of the judiciary, by the discovery of incriminating evidence."); Jonathon D. Casper, Kennette Benedict, and Jo L. Perry, *The Tort Remedy in Search and Seizure Cases: A Case Study in Juror Decision Making*, 13 LAW & SOC. INQUIRY 279 (1988) (reporting results of simulated jury study that found knowledge of the outcome of the search affected assessment of the legality of the search); Myron W. Orfield, *The Exclusionary Rule in Chicago's Courts*, 63 COLO. L. REV. (1991) (forthcoming).

For more general treatments of the tendency for factual evidence of guilt to overcome procedural standards in criminal cases, see Francis A. Allen, *A Serendipitous Trek Through the Advance-Sheet Jungle: Criminal Justice in Courts of Review*, 70 IOWA L. REV. 311, 316-17 (1984) (appellate decisions in criminal cases reflect popular pressure to be tough on crime); Thomas Y. Davies, *Affirmed: A Study of Criminal Appeals and Decision Making Norms in a California Court of Appeal*, 1982 AM. B. FOUND. RES. J. 543, 625-32 (substantive justice affects appellate review of legal issues); *see also* DOREEN J. MCBARNET, CONVICTION: LAW, THE STATE AND THE CONSTRUCTION OF JUSTICE 158 (1981); Thomas Y. Davies, *Do Criminal Due Process Principles Make a Difference?*, 1982 AM. B. FOUND. RES. J. 247.

Numerous commentators have expressed concern about the underenforcement of Fourth Amendment standards by trial courts. *See* Yale Kamisar, *The Warren Court (Was It Really So Defense-Minded?), the Burger Court (Is It Really So Prosecution-Oriented?), and Police Investigatory Practices, in* THE BURGER COURT: THE COUNTER-REVOLUTION THAT WASN'T 82 n.122 (Vincent Blasi ed., 1983) ("[M]ost front-line courts limit application of the exclusionary rule to willful police illegality; official adoption of good-faith exception may reduce the actual operation of the rule to the vanishing point."); Anthony G. Amsterdam, *The Supreme Court and the Rights of Suspects in Criminal Cases*, 45 N.Y.U. L. REV. 785, 785-93 (1970); John Kaplan, *The Limits of the Exclusionary Rule*, 26 STAN. L. REV. 1027, 1045 (1974).

Numerous critics of the exclusionary rule have contended trial judges tend to bend search rules to admit incriminating evidence. *See* Gerald G. Ashdown, *Good Faith, the Exclusionary Remedy, and Rule-Oriented Adjudication in the Criminal Process*, 24 WM. & MARY L. REV. 335, 383-84 (1983) (judges would take a broader view of Fourth Amendment rights if there were a good faith exception because suppression could still be avoided); Frederick A. Bernardi, *The Exclusionary Rule: Is a Good Faith Standard Needed to Preserve a Liberal Interpretation of the Fourth Amendment?*, 30 DEPAUL L. REV. 51 (1980); Philip S. Coe, *The ALI Substantiality Test: A Flexible Approach to the Exclusionary Sanction*, 10 GA. L. REV. 1, 30 n.190 (1975); Dallin H. Oaks, *Studying the Exclusionary Rule in Search and Seizure*, 37 U. CHI. L. REV. 665, 747 (1970); William A. Schroeder, *Deterring Fourth Amendment Violations: Alternatives to the Exclusionary Rule*, 69 GEO. L. J. 1361, 1412-21 (1981); Malcolm R. Wilkey, *A Call for Alternatives to the Exclusionary Rule: Let Congress and the Trial Courts Speak*, 62 JUDICATURE 351, 356 (1979) ("[T]rial judges are blatantly hypocritical in construing [search standards] because . . . the illogical penalty

are too flexible thus reduce the likelihood that citizens' privacy will be protected.

This article does not provide a comprehensive demonstration that the Rehnquist Court too often disregards doctrine, though such a demonstration could be compiled given adequate time and paper. Instead, I endeavor to show the extent of the Court's willingness to disregard settled doctrine through a close analysis of the three key doctrinal claims that constitute the rationale for a single Rehnquist Court decision that adopts a weak search standard, *Illinois v. Rodriguez*[10] (a case in which I appeared "of counsel" for the defendant[11]).

The question addressed in *Rodriguez* is whether a warrantless police entry of a home can be based on third-party consent if the third party is not a resident of the home and does not possess any authority over the home.[12] Predictably, the Court held "seeming consent"[13] is enough to satisfy the Fourth Amendment.[14] Justice

---

of total exclusion . . . is damaging to the cause of justice."); *see also* United States v. Leon, 468 U.S. 897, 925 n.26 (1984), quoting Schroeder, *supra* (exception to the exclusionary rule will make it less tempting for judges to bend Fourth Amendment standards to avoid releasing criminal); Allen v. McCurry, 449 U.S. 90, 115 (1980) (Blackmun, J., dissenting) ("A trial court, faced with the decision whether to exclude relevant evidence, confronts institutional pressures that may cause it to give a different shape to the Fourth Amendment right from what would result in civil litigation of a damages claim [because] a trial court, at least subconsciously, must weigh the potential damage to the truth-seeking process caused by excluding relevant evidence."); Brief of the United States at 76, United States v. Leon, 468 U.S. 897 (1984) (there is an "obvious reluctance of judges to condemn questionable practices under the Fourth Amendment when they know the result of their decision will be the freeing of a guilty defendant"); Amicus Brief of the United States at 55, Illinois v. Gates, 462 U.S. 213 (1983) (same).

But see United States v. Leon, 468 U.S. 897, 916-17 (1984) ("There exists no evidence suggesting that judges or magistrates are inclined to ignore or subvert the fourth amendment . . . .").

10. 110 S. Ct. 2793 (1990).

11. I appeared of counsel on a pro bono basis at the invitation of James W. Reilley of DesPlaines, Illinois, counsel for the respondent Edward Rodriguez. I previously had the privilege of assisting Mr. Reilley in a similar fashion in the reargument in Illinois v. Gates, 462 U.S. 213 (1983).

12. United States v. Matlock, 415 U.S. 164, 171 n.7 (1974) (articulating a co-inhabitant standard for common authority to consent). *See infra* notes 53-59 and accompanying text.

13. 110 S. Ct. at 2800 n.*. I adopt Justice Scalia's phrase "seeming consent" as a shorthand reference to a situation in which there is a reasonable appearance that a "consenting" third party posesses authority to consent even though the third party does not possess such authority in fact. I do not refer to the *Rodriguez* standard as an apparent authority to consent standard because that term creates confusion with the unrelated agency law concept of apparent authority. *See infra* note 87. Similarly, I do not refer to the *Rodriguez* standard as a reasonable police belief

Scalia's opinion for the six Justice majority adopted a weak reasonable-appearance-of-authority-to-consent standard by holding a police intrusion of a residence in which "consent"[15] is obtained from a third party complies with the Fourth Amendment if "the facts available to the officer at the moment . . . [would] 'warrant a man of reasonable caution in the belief' that the consenting party had authority over the premises," regardless of whether the third party actually possessed such authority.[16]

Rodriguez may not appear at first to be a particularly important Fourth Amendment decision. The majority opinion is so artfully written in a low-key style that it does not appear to present much controversy. As a practical matter, moreover, it may not seem the decision will make much difference. Although consent is probably the most common ground offered to establish the constitutionality of police intrusions,[17] it does not appear that the specific situation

---

standard because that misstates the standard Justice Scalia articulates. *See infra* note 30.

14. Professor LaFave predicted the Court would decide the appearance of authority to consent issue as it did because some lower courts had adopted that approach and because he believed the Court was not inclined to disfavor consent intrusions. WAYNE R. LAFAVE, SEARCH AND SEIZURE: A TREATISE ON THE FOURTH AMENDMENT (2d ed.) § 8.3(g) 266 (1987). *See infra* note 84.

15. One persistent source of potential confusion is that the word "consent" is used in two distinct ways. First, it is used merely to indicate the act of inviting or agreeing to an intrusion. Second, it is also used to indicate the legal condition of agreement to an intrusion—the act of agreement plus the authority to give it. The first meaning is a necessary but insufficient condition for the second meaning. To keep these two usages separate, I put the words "consent," "consented" or "consenting" in quotation marks where it refers only to an act of agreement. I use consent without quotation marks where it is used to denote the legal condition of agreement. I do not alter the treatment of the word when it appears in quotations.

16. 110 S. Ct. at 2801 (quoting Terry v. Ohio, 392 U.S. 1, 21-22 (1968)). Note Justice Scalia speaks of the "consenting party" even though the person did not necessarily possess authority to consent. *See supra* note 15.

17. One study reports, "The vast majority of searches are conducted without a warrant, usually with the consent of the suspect (or someone in legal control of the area to be searched) or incident to the arrest of the suspect." RICHARD VAN DUIZEND, L. PAUL SUTTON AND CHARLOTTE A. CARTER, THE SEARCH WARRANT PROCESS: PRECONCEPTIONS, PERCEPTIONS, PRACTICES 19 (1985). One detective estimated as many as 98% of the searches were by consent, but other estimates were as low as 10%. *Id.* at 19, 68. The same study reports police officers stated consent searches "nearly always stood up under challenge in court," even though a number of judges "expressed uncertainty over the degree to which consents to search were truly voluntary." *Id.* at 69-70.

Consent is invoked to justify police searches in a wide variety of settings including searches of airline passengers, United States v. Mendenhall, 446 U.S. 544 (1980), automobiles, Florida v. Jimeno, 111 S. Ct. 1801 (1991), bus passengers, Florida v. Bostick, 111 S. Ct. 2382 (1991), and residences, United States v. Matlock, 415 U.S. 164 (1973). It is plausible that consent would be the reason offered for most

addressed in *Rodriguez*—in which the "consenting" third party had *no authority at all* regarding the premises—has arisen with much frequency.[18] It is certainly possible, however, and perhaps even likely such situations will appear more frequently in the future.[19]

Regardless, the majority rationale in *Rodriguez* is worthy of close attention because the manipulations and evasions of existing doctrine in the opinion reveal the degree to which the majority Justices are willing to ignore well-settled precedents, concepts, and principles to uphold a highly invasive police intrusion into a citizen's home despite the fact that the case does not necessarily have broad implications for efficacious law enforcement practices.[20]

The rationale for *Rodriguez* is constructed from three necessary doctrinal claims. First, "consent" need not be based on the conduct of a person who actually possesses authority over the premises involved. Second, "consent" is one of various elements that can satisfy the Fourth Amendment's reasonableness standard for police intrusions. Third, any error the police made about a factual matter comports with the Fourth Amendment provided the mistake was understandable in the circumstances. I argue that each of these three claims amounts to an unacknowledged distortion of prior doctrine.

A. Overview of This Article

In Part II, I provide background information about the police intrusion into Rodriguez's apartment and outline the procedural development of the issues in *Rodriguez*. In Part III, I briefly map out

---

warrantless police intrusions of residences because there are few alternative legal justifications for warrantless intrusions in that setting. An unknown proportion of consent intrusions are justified on the ground that the police obtained consent from a third party—someone other than the target of the intrusion.

18. There has been a substantial volume of litigation reported regarding third-party authority issues in "consent" intrusions. *See, e.g.*, LAFAVE *supra* note 14, § 8.3 at 236. Relatively few of the reported cases, however, involve the threshold issue addressed in *Rodriguez* of whether a "consenting party" possesses *any* authority over the premises. *See, e.g.*, United States v. Matlock, 415 U.S. 164 (1974). Many of the reported cases raise issues regarding the *scope* of a consenting third-party's authority with regard to specific rooms or containers in a residence over which the third party has authority. *See, e.g.*, Frazier v. Cupp, 394 U.S. 731 (1969) (a co-user of a duffle bag has authority to consent to a search of all pockets of the bag regardless of defendant's claim co-user was not allowed to use the pocket in which the incriminating evidence was found), discussed *infra* note 124.

19. The Court's announcement that "seemingly consented searches" are acceptable may encourage police officers to be less hesitant to act on the basis of dubious third-party "consent" in the future. *See infra* text accompanying notes 278-87.

20. This is not to say the *Rodriguez* majority's disregard of doctrine is immediately apparent. For a neutral description of *Rodriguez* that does not raise any of the criticisms I make, see LAFAVE, *supra* note 14, at 1991 Supp. § 8.3, 36-39.

the rationale Justice Scalia presents in the majority opinion. I then address the three key doctrinal claims that form the basis for the majority's rationale.

In Part IV, I examine the majority's unprecedented and logically contradictory treatment of the concept of consent. Justice Scalia characterizes consent as though it were one of various elements that make a search "reasonable" and therefore satisfy the Fourth Amendment. He asserts authority to consent, like other such elements, must be evaluated from an *ex ante* police perspective rather than in terms of authority in fact. He further asserts the important question is not whether consent was actually given, but whether the police officers could have believed it had been given.

There is no precedent for this notion of "consent." Before *Rodriguez*, the Court consistently viewed consent as though it operated as a waiver or release of the consenting person's *own* right. Under this understanding of consent, it is axiomatic that the inquiry begins and ends with the question of whether a person who actually had authority to consent voluntarily gave consent. The police intrusion of Rodriguez's apartment was not consented to within the traditional understanding of that term. Justice Scalia's majority opinion evades that outcome, however, through the simple expedient of ignoring prior doctrine.

Significantly, although *Rodriguez* rests on premises that are inconsistent with the traditional understanding of consent, it does not actually repudiate the traditional concept of consent and replace it with "seeming consent." Instead, *Rodriguez* adopts an either/or formulation. "Seeming consent" is the standard to be applied to assess "consent," unless the facts surrounding the intrusion are such that it can only be upheld under an authority in fact standard, in which case the latter standard should be applied. *Rodriguez*'s treatment of consent, in short, is blatantly driven by the desire to reach a progovernment result.

In Part V, I explore the downgraded understanding of Fourth Amendment reasonableness implicit in Justice Scalia's claim that consent is one of various elements that can make a warrantless police intrusion of a home "reasonable."[21] This characterization implies it is "reasonable" for the police to enter a residence once they have reason to think consent has been given.[22] This claim, however, does

---

21. To avoid unnecessary confusion between Justice Scalia's use of the words "reasonable" and "reasonableness" and the traditional concept of Fourth Amendment reasonableness, I put the words "reasonable" and "reasonableness" in quotation marks when they are used in the loose, colloquialized way in which Justice Scalia employs them. *See infra* notes 252-54 and accompanying text. I omit the quotation marks when they are employed in a way that is consistent with the traditional understanding of reasonableness in the Fourth Amendment as discussed *infra* notes 184-97 and accompanying text.

22. The *Rodriguez* majority opinion does not make this claim in so many

not comport with the traditional concept of Fourth Amendment reasonableness, especially as that concept has been applied to searches of homes.

Under the traditional understanding of Fourth Amendment reasonableness, the core requirement for a constitutional intrusion of a citizen's residence by police is probable cause.[23] Cause for a search must be based on information that creates a substantial expectation that a particular person or thing will be found at a particular location. Even validly given consent cannot provide cause for an intrusion;[24] it merely provides permission for one. For this reason, prior decisions determined validly consented intrusions to be constitutional only because consent withdrew a resident's expectation of privacy and thus made the reasonableness requirement of the Fourth Amendment *inapplicable* to the resulting police intrusion.

Thus, when *Rodriguez* claims a "seemingly consented search" *satisfies* the Fourth Amendment, it employs a loose, colloquialized notion of "reasonableness" (understandableness under the circumstances[25]) that omits any cause requirement. *Rodriguez* drains the operative meaning from the traditional concept of Fourth Amendment reasonableness. It trivializes the core requirement of the Fourth Amendment in the context of an extremely invasive police intrusion of a home.

In Part VI, I examine the *Rodriguez* majority's claim that a police assessment of a third-party's authority to consent is a factual determination that falls under a general rule that understandable police errors about facts constitute compliance with the Fourth Amendment. There are two defects in this claim. First, the majority's characterization of the police error in *Rodriguez* as factual is arbitrary. A police error regarding authority to consent could occur because of factual confusion about the residence of a consenting third party. It also could occur, however, because of police ignorance of the applicable legal standard for such authority.[26]

---

words. Chief Justice Rehnquist, however, recently cited *Rodriguez* as authority for this proposition in a majority opinion joined by all of the Justices in the *Rodriguez* majority. *See* Florida v. Jimeno, 111 S. Ct. 1801, 1803 (1991). *See also infra* notes 149-50 and accompanying text.

23. *See infra* notes 223-25 and accompanying text. There are some police intrusions in which the Court has held a lower showing of cause than probable cause will suffice because they are deemed to involve only minimal intrusions of citizens' privacy. *See* cases cited *infra* notes 206-13. Before *Rodriguez*, the Court required a police intrusion into a residence to be based on a showing of probable cause even if the warrant requirement was inapplicable.

24. Probable cause consists of information giving rise to a particularized expectation about what an intrusion will produce. Police are not required, however, to have a particularized expectation to request consent for an intrusion. *See infra* notes 258-64 and accompanying text.

25. 110 S. Ct. at 2800 (quoting Hill v. California, 401 U.S. 797, 803-04 (1971)).

26. The applicable legal standard is the co-inhabitant standard set out in

The record in *Rodriguez* does not contain any finding regarding the nature of the police error. Certain aspects of the record, however, provide strong circumstantial evidence that the intrusion occurred because of police ignorance of the legal standard for authority to consent, not because of factual confusion over the residence of the third party. The nature of the police error is important because there is no way to describe police ignorance of a legal search standard as "understandable" or "reasonable."

Second, even assuming the police error was factual, the general rule asserted in *Rodriguez* substantially overstates the excusability of factual police mistakes. The Court has excused understandable police errors about facts in the past only in connection with facts that run to the assessment of the probable cause component of Fourth Amendment reasonableness or to facts that run to whether there is an exigency that excuses the warrant requirement.[27] In that context, the allowance of understandable factual errors is not controversial: the probable cause standard is probabilistic by definition, so an understandable mistake does not transgress that legal standard. The police mistake discussed in *Rodriguez*, however, is not about probable cause; it is about the validity of a third-party's legal authority. Legal authority is not a probabilistic standard, and the Court previously has treated police errors about legal authority as inherently unreasonable under the Fourth Amendment.[28] The *Rodriguez* majority's willingness to excuse a police error about legal authority demonstrates a disregard for enforcing legal search standards.

Therefore, each of the three critical claims from which the *Rodriguez* rationale is constructed violates settled legal principles and constitutional doctrine. Each was manufactured on the spot in order

---

United States v. Matlock, 415 U.S. 164 (1974). *See infra* notes 53-58 and accompanying text.

27. *See* Maryland v. Garrison, 480 U.S. 79 (1987); Hill v. California, 401 U.S. 797 (1971); Brinegar v. United States, 338 U.S. 160 (1949). *Brinegar* and *Hill* explicate the probable cause standard. *See infra* notes 314-26 and accompanying text. Justice Stevens's majority opinion in *Garrison* is not explicitly argued in terms of the probable cause standard. The police error in *Garrison*, however, is an error running to probable cause, so Justice Stevens's discussion of the excusability of "reasonable" police errors should be understood in that light. *See infra* notes 328-61 and accompanying text.

Police errors regarding exigency are discussed *infra* notes 372-73.

28. The Court implicitly recognized the nonprobabilistic nature of legal authority in United States v. Leon, 468 U.S. 897 (1984). In *Leon* the Court created an exception to the exclusionary rule for evidence seized when police officers make searches in objectively reasonable reliance on constitutionally defective search warrants. *Id.* at 922. *Leon* does not suggest the "reasonableness" of the police error in that setting can make the search constitutional. To the contrary, it creates an exception to the exclusionary rule precisely because it recognizes the search was unconstitutional regardless of the "reasonableness" of the police error. The Court's characterization of the doctrinal significance of an understandable police error about legal authority in *Rodriguez* is a drastic departure from the understanding expressed in *Leon*. *See infra* notes 363-67 and accompanying text.

to reach the ideologically desired result. The same also can be said of the objective standard for "seeming consent" set out in *Rodriguez*.[29] As I explain in another article, the majority opinion formulates this standard in such a way as to evade the potentially important inquiries in assessing the "reasonableness" (understandableness) of police errors regarding authority to consent.[30]

---

29. *See infra* text accompanying note 83.

30. *See* Thomas Y. Davies, *A Search Standard that Does Not Measure Up: How the Rehnquist Court Made it Too Easy to Find "Seeming Consent" in* Rodriguez (forthcoming). I argue the standard announced in *Rodriguez* for assessing "seeming consent" has virtually no operative content. First, the language in which the standard is stated ("warranted in the belief") is adopted from the formulation for the reasonable suspicion standard adopted for less invasive intrusions in Terry v. Ohio, 392 U.S. 1 (1968). See *infra* note 83 and accompanying text. In light of that, the *Rodriguez* threshold for finding a sufficiently "reasonable" appearance that a third party possesses authority to consent appears to be set at whether the police have enough information to form a suspicion that the third party has such authority. This is a minimal, "any information" standard on its face.

Second, *Rodriguez* does not appear to require the police to show they actually believed the third party had authority to consent. *Rodriguez* describes the standard for "seeming consent" as an "objective standard." *See infra* text accompanying note 83. Other recent Supreme Court opinions that employ "objective reasonableness" standards in Fourth Amendment cases forbid any assessment of the actual subjective intentions or beliefs of the searching officers and allow assessments only as to how a hypothetical reasonable officer would respond to the situation. *See, e.g.*, United States v. Leon, 468 U.S. at 922 n.7 (assessment of good faith and objective reasonableness not permitted to involve "expedition into the minds of the police officers"). There is every reason to assume the Court also intends the *Rodriguez* standard to be applied only hypothetically. Consequently, *Rodriguez* does not adopt a police "reasonable belief" standard.

Third, *Rodriguez* finesses the basic question of whether the officers at the scene knew and applied the *Matlock* legal standard for assessing a third party's authority. Hypothetically reasonable officers always know the law, but in reality officers are sometimes ignorant. The record in *Rodriguez* suggests the police were actually ignorant of the co-inhabitant legal standard for third-party consent, not merely misinformed as to where Fischer resided. *See infra* notes 302-08 and accompanying text. The hypothetical formulation in *Rodriguez* allows police mistakes about legal standards (which lead the police to ask the wrong questions) to be excused as though they were factual errors (that is, as though the police received the wrong answers).

Fourth, and perhaps most important, the *Rodriguez* majority declined to impose any duty of investigation on the police with regard to assessing the third-party's co-inhabitant status. The *Rodriguez* opinion treats the police as though they were merely passive recipients of windfall information available at the moment instead of active gatherers of information who can readily increase the information at hand simply by asking additional questions of the consenting party.

Fifth, the absence of any statement of a police duty to investigate is especially striking in the context of the highly atypical circumstances under which the third-party "consent" was obtained in *Rodriguez*. *See infra* note 38. However, there is no mention of whether or how such circumstances are to be assessed under the standard announced in *Rodriguez*.

In sum, the only concern evident in the extremely cursory discussion of the objective standard in *Rodriguez* is the majority's insistence on making it as easy as possible

I conclude this article by briefly summing up the implications of the Rehnquist Court's disregard of Fourth Amendment doctrine in *Rodriguez*. The majority opinion in *Rodriguez* reflects the majority Justices' disinclination to enforce the Fourth Amendment unless the government so blatantly abuses citizens' privacy that no excuse for the government misconduct can be concocted. I neither hold out a silver lining nor suggest any program for remedying the currently downgraded state of Fourth Amendment rights. There is no realistic possibility that the present majority Justices will be persuaded to take Fourth Amendment doctrine seriously. Neither is there any likelihood that the executive or legislative branches will champion the enforcement of Fourth Amendment rights. My purpose in writing is simply to offer a lucid assessment of the dismal state of what now passes for Fourth Amendment doctrine.

II. THE FACTUAL BACKGROUND AND PROCEDURAL HISTORY OF *RODRIGUEZ*

A. The Police Intrusion

On the morning of July 26, 1985, Gail Fischer visited her boyfriend, Edward Rodriguez, at his apartment in Chicago. Fischer, who previously had lived with Rodriguez,[31] visited him frequently at his apartment. During that particular visit they argued, and Rodriguez beat Fischer. Fischer then went back to her mother's home, where she was living with her two children. Her mother called the Chicago Police Department and reported the beating. Three Chicago police officers interviewed Fischer at her mother's home,[32] and Fischer told them she wanted Rodriguez arrested. She also told them Rodriguez probably was sleeping in the bedroom of his apartment. At least one of the police officers recognized Rodriguez's name as a suspected drug dealer.[33]

Justice Scalia's opinion reports "It is unclear whether [Fischer] indicated [to the police] that she currently lived at the apartment, or

---

for police to invoke, and for reviewing courts to uphold, claims of "seemingly consented searches."

31. Fischer and her two children lived with Rodriguez until about a month before the incident, when they moved in with her mother. There is some indication Fischer expected to move back in with Rodriguez after toilet training her children.

32. Supplemental Record at 13-14. Initially, one patrol officer, Officer Tenza, responded to the complaint. He subsequently called in Officers Entress and Guitierez from a tactical unit.

33. *Rodriguez* Joint Appendix at 22. Officer Entress testified he asked Fischer if Rodriguez had drugs but Fischer did not answer the question. Joint Appendix 23. Fischer's mother testified Fischer did tell the police that Rodriguez had drugs. Joint Appendix at 49.

only that she used to live there."[34] The reason this is unclear, which Justice Scalia does not mention, is because the police testimony regarding what Fischer said about where she lived is inconsistent. At the preliminary hearing (held six weeks after the intrusion) Officer Entress (the only officer who testified) said Fischer "stated *she used to live*" at Rodriguez's apartment.[35] This statement conclusively shows Fischer did not have legal authority to consent on the date of the intrusion. At the suppression hearing (held a full year after the intrusion) Entress changed his testimony and said Fischer "stated to me she *had been living*" in the apartment.[36] Fischer testified she never told the officers she was living with Rodriguez.[37]

There is no dispute that the officers did not ask Fischer many questions about her living arrangements. Although the factual setting was highly atypical for third-party consent,[38] Entress testified he did not "go into specifics" by asking Fischer any further questions about her residence.[39] Entress did testify Fischer had referred to Rodriguez's apartment as "our" apartment,[40] that she had said she had, and did have, a key to the apartment (though he did not testify he had asked for, or she had offered, any explanation of how she came to have

---

34. *Rodriguez*, 110 S. Ct. at 2797.

35. Supplemental Record at 16 (emphasis added). Entress also conceded his memory of her statement may have been better when he testified at the preliminary hearing than at the later suppression hearing. Joint Appendix at 12. Entress's original testimony suggests the real mistake made by the police was their unfamiliarity with the *Matlock* co-inhabitant standard. *See infra* notes 304-08 and accompanying text.

36. Joint Appendix at 10. Entress testified Fischer's "exact words" were that "she had been living there." This cannot be literally true; Fischer would not have said "she." It is also difficult to believe she would have said "I had been living there." This second version of the statement Entress attributed to Fischer—"she had been living there"—is also grammatically ambiguous as to whether Fischer was living there on the date of the police intrusion.

37. Joint Appendix at 65.

38. Reported descriptions of third-party consent intrusions of residences typically exhibit a cluster of features: the police go to a residence with a desire to arrest a suspect or make a search; the suspect is absent but a spouse or other adult family member is present at the residence; and this third person, who exhibits no hostility toward the suspect, gives consent to the police to enter or search. The police entry of Edward Rodriguez's apartment deviated from virtually every feature of this typical pattern.

First, the police did not believe the suspect, Rodriguez, was absent from the residence, but understood he was sleeping inside the apartment at the time. Second, Gail Fischer, the third-party, was not Rodriguez's wife, family member, or relation, but merely his girlfriend. Third, Fischer was not present at the apartment the police wanted to enter; they found her at her mother's residence (where she was then living with her children) and took her to Rodriguez's apartment. Fourth, Fischer was clearly hostile toward Rodriguez; she wanted him arrested for battery.

39. Joint Appendix at 10. Entress testified he "didn't go into specifics with [Fischer] as if she had just moved out or anything like that." *Id.*

40. Joint Appendix at 26-27.

it),[41] and that she had said she had possessions and clothing in the apartment.[42]

The three officers did not apply for a battery arrest warrant,[43] although they would have been able to show probable cause,[44] and there was no exigency that prevented them from applying for one.[45] They also did not attempt to develop probable cause for a search warrant for drugs. Instead, the police chose to enter Rodriguez's apartment on the basis of Fischer's "consent." The officers drove Fischer and her mother to Rodriguez's apartment, and Fischer opened the door with her key. The police then entered the apartment—without knocking or alerting Rodriguez to their presence—and spent at least a minute and a half inside prior to waking and arresting Rodriguez,[46] who was sleeping in his bedroom as they expected.[47]

---

41. Joint Appendix at 6, 10, 16, 26.

42. The testimony regarding what Fischer said about her possessions and clothing is in conflict. Officer Entress testified Fischer said "all her clothing" and other possessions were at the apartment on the date of the intrusion. Joint Appendix at 6, 25. Fischer and her mother testified that Fischer had taken her clothing as well as her children's clothing with her when she moved out. *Id.* at 38-40, 63. It is difficult to imagine any motive Fischer could have had to misstate this information to the police at the time of her interview. The suppression judge found that Fischer had taken her clothes to her mother's. *Id.* at 96.

The fact that any discussion occurred regarding the whereabouts of Fischer's clothing or possessions is significant because it casts doubt on the characterization that the police actually thought Fischer lived in Rodriguez's apartment. *See infra* note 307 and accompanying text.

43. The police passed up several other options. Under Payton v. New York, 445 U.S. 573 (1980), police may not make a forcible entry of a residence for the purpose of effecting an arrest of a person inside unless they have obtained an arrest warrant. *Payton*, however, does not prohibit the officers from seeking the consent of a resident to enter the apartment for the purpose of making an arrest of the resident. It is not clear whether *Payton* prevents the officers from simply knocking on the door and then arresting the resident when he opens it. *Cf.* United States v. Santana, 427 U.S. 38 (1976).

44. Fischer's statements to the officers regarding the beating and the officers' observation of her injuries would have been adequate to show probable cause for an arrest warrant. The officers' ability to show probable cause is not incorporated into the rationale or holding in *Rodriguez*, however. *See infra* notes 268-70 and accompanying text.

45. Although Fischer had been injured, she was not in any continuing or immediate danger; several hours had passed since the beating and she was in the company of her mother, who was a Cook County deputy sheriff, and the officers. Fischer was not so traumatized, moreover, that the police sought medical assistance. In fact, they transported her back to the scene of the beating so she could open the door with the key she had.

46. Supplemental Record at 18-20.

47. *Id.* The police could not have entered Rodriguez's apartment with such stealth if they had chosen to proceed with an arrest warrant or a search warrant, because there were no grounds for a no-knock warrant. The fact the police were interested in the possibility that Rodriguez might have drugs, see *supra* note 33 and

While in the apartment, the police allegedly saw drug paraphernalia and an open Tupperware container containing white powder that appeared to be cocaine in plain view in the living room, as well as two open briefcases that contained plastic bags containing a similar white powder in plain view in the bedroom. The police seized the suspected drugs and Rodriguez was prosecuted for possession of a controlled substance with intent to deliver.[48]

B. Proceedings Regarding Suppression of the Seized Drugs

Rodriguez's attorneys moved for suppression of the drugs seized in Rodriguez's apartment on the basis that the seizure was unconstitutional. Two settled legal principles shaped this issue. First, because the discovery of the drugs was a fruit (a direct result) of the police entry of the apartment,[49] regardless of the allegedly plain view character of the discovery,[50] the constitutionality of the seizure depended on the constitutionality of the police entry. Second, *Payton v. New York*[51] teaches that, in the absence of consent or exigent circumstances,

---

accompanying text, strongly suggests the officers chose to enter on the basis of Fischer's "consent" because of the opportunity to enter and look around without alerting Rodriguez. *See infra* notes 281-82 and accompanying text.

48. 110 S. Ct. at 2797. Fischer signed a battery complaint after Rodriguez's arrest, but no battery charge was filed.

49. Under the fruit of the poisonous tree doctrine, a seizure or confession that occurs as a consequence of a constitutional violation is also unconstitutional. *See* Taylor v. Alabama, 457 U.S. 687 (1982); Dunway v. New York, 442 U.S. 200 (1979); Brown v. Illinois, 422 U.S. 590 (1975); Nardone v. United States, 308 U.S. 338 (1939); Silverthorne Lumber Co. v. United States, 251 U.S. 385 (1920). If the police entry of Rodriguez's apartment violated the constitution, then the police discovery and seizure of the drugs were unconstitutional regardless of whether or not the drugs were in plain view once the officers entered.

50. The prosecution sought to justify the discovery and seizure of the drugs as distinct from the police entry of the apartment by invoking the plain view doctrine. Under the plain view doctrine, no protected privacy interest is violated if a police officer seizes items that he sees from a place where he is legally entitled to be and that he has probable cause to believe are evidence or contraband. *See, e.g.*, Harris v. United States, 390 U.S. 234, 236 (1968) ("It has long been settled that objects falling in the plain view of an officer who has a right to be in a position to have that view are subject to seizure and may be introduced into evidence."). The Court recently clarified that there is no requirement that an in plain view discovery must be inadvertent. *See* Horton v. California, 110 S. Ct. 2301 (1990).

The plain view allegations regarding the drugs seized in the *Rodriguez* intrusion are important because Fischer is alleged to have "consented" only to the police entry for the purpose of arresting Rodriguez; there is no claim that Fischer "consented" to a search of the apartment. Therefore any seizure of evidence could be justified only on the grounds the evidence was discovered in plain view or in the course of a search incident to the arrest. The scope of a search incident to arrest, however, is limited to the area in the immediate control of the arrestee. Chimel v. California, 395 U.S. 752 (1969).

51. 445 U.S. 573 (1980) (police may not enter a residence to make an arrest

police may not enter a residence to make an arrest unless they first obtain a warrant. As noted above, there was no exigency. Hence, the warrantless police entry of Rodriguez's apartment violated his Fourth Amendment right unless Fischer's "seeming consent" provided a justification for the entry.

Rodriguez's attorneys argued Fischer's alleged "consent" to the police entry was immaterial[52] because she did not possess the requisite authority under *United States v. Matlock*[53] to give valid and effective consent.[54] *Matlock*, the leading authority on third-party consent,[55] stated a third party possesses authority to consent to an intrusion of a residence only if the third party is a "co-inhabitant"[56] who "possess[es] common authority" over the residence.[57] *Matlock* further defined common authority as "mutual use of the property by persons generally having joint access and control for most purposes."[58] *Matlock* thus sets out a threshold test: a third party has no authority to

---

of a resident, in the absence of a genuine exigency or consent, unless the police first obtain an arrest warrant). *See also* Minnesota v. Olson, 495 U.S. 91 (1990) (police may not enter a residence to arrest overnight guest unless the police first obtain an arrest warrant).

52. Rodriguez's attorneys also argued Fisher's "consent" was not voluntary. Under Schneckloth v. Bustamonte, consent must be voluntary; it is ineffective if it is given as a result of express or implied duress or coercion. 412 U.S. 218, 248 (1973). There is conflicting testimony as to whether Fischer voluntarily "consented." Officer Entress testified Fischer volunteered to let the police into the apartment. Joint Appendix at 6. Fischer, however, testified the police officers told her she "had to" let them into the apartment or Rodriguez could not be arrested for beating her. *Id.* at 77-78. The resolution of this factual dispute holds implications for the assessment of the validity of Fischer's "consent" because police deception is generally treated as a form of coercion. *See* Bumper v. North Carolina, 391 U.S. 543 (1968). The judge presiding over the Rodriguez suppression hearing did not rule on this argument.

53. 415 U.S. 164 (1974). For commentary regarding *Matlock* and third-party consent, see John B. Wefing & John G. Miles, Jr., *Consent Searches and the Fourth Amendment*, 5 SETON HALL L. REV. 211 (1974); James B. White, *The Fourth Amendment as a Way of Talking About People: A Study of Robinson and Matlock*, 1974 SUP. CT. REV. 165, 216-32; Robert Dechene, Note, *The Problem of Third-Party Consent in Fourth Amendment Searches: Toward a "Conservative" Reading of the* Matlock *Decision*, 42 MAINE L. REV. 159 (1990); *see also* LAFAVE, *supra* note 14 at § 8.3.

54. The terms "effective consent" and "valid consent" are technically redundant; either there is consent or there is not. I will sometimes use these terms, however, to prevent confusion with the use of the word "consent" to refer only to the act of inviting or permitting, as opposed to the legal condition of consent. *See supra* note 15.

55. "Third party" is a common appellation for any "consenting party" other than the person who is the target of the search or arrest. *See, e.g.*, United States v. Matlock, 415 U.S. 164, 166 (1974).

56. *Id.* at 172 n.7. *Matlock*, however, leaves open a possibility that some "other sufficient relationship to the premises or effects sought to be inspected" might suffice for common authority. *Id.*

57. *Id.* at 171.

58. *Id.* at 172 n.7.

consent to an intrusion of a residence unless the third party is a co-inhabitant who exercises common control.[59]

Rodriguez's attorneys argued Fischer was not a co-inhabitant and did not possess authority to consent under the *Matlock* standard. The prosecution argued Fischer did meet the *Matlock* standard.[60] The judge presiding over the suppression hearing ruled Fischer did not possess authority to consent to the police entry under *Matlock* because she "was not a usual resident" of the apartment but was "a rather infrequent visitor or resident or guest or invitee."[61]

The prosecution also offered an alternative argument for upholding the police entry of the apartment; namely, "that, even if Fischer did not possess common authority over the premises, there was no Fourth Amendment violation if the police reasonably believed at the time of their entry that Fischer possessed the authority to consent."[62] The prosecution premised its argument on a footnote in *Matlock* that noted the Court had not reached a question raised by the government in that case: whether a reasonable although erroneous belief held by searching officers that a third party had authority to consent would be enough to satisfy the Fourth Amendment.[63]

The suppression judge rejected the prosecution's alternative argument because Illinois case law consistently required actual authority for consent.[64] He ordered the drugs suppressed without making any findings of fact regarding the information the police had at the time or the character of their judgment and conduct.[65] The Illinois Appel-

---

59. *Rodriguez*, 110 S. Ct. at 2797. The Court recently clarified that the standard for authority to consent to an intrusion of a residence is higher than the standard for standing to challenge an intrusion of a residence. *See* Minnesota v. Olson, 110 S. Ct. 1684, 1689-90 (1990). Only a co-inhabitant possesses authority to consent to an intrusion of a residence, but an overnight guest may have standing to challenge a search. *Id.* The relationship between the rulings in *Rodriguez* and *Olson* is discussed *infra* note 172 and accompanying text.

60. The prosecution argued the police reasonably concluded Fischer had authority over Rodriguez's apartment given the following facts: Fischer said she "had been living" in the apartment; she referred to the apartment as "our" apartment; she possessed a key; and she said she had possessions in the apartment.

61. Joint Appendix at 95, 96.

62. *Rodrigeuz*, 110 S. Ct. at 2797. The standard articulated in *Rodriguez* does not require the police to actually have a reasonable belief regarding the third party's authority. *See supra* note 30 and accompanying text.

63. The *Matlock* footnote is discussed *infra* notes 151-54 and accompanying text.

64. 110 S. Ct. at 2797-98. Illinois was one of a number of states that declined to recognize the apparent authority to consent doctrine. *See infra* note 84 and accompanying text. Therefore, Illinois decisions required actual authority for third-party consent. *See* People v. Vought, 528 N.E.2d 1095, 1110 (1988) (referring to decisions in prior Illinois cases rejecting the apparent authority to consent doctrine). Justice Scalia's majority opinion concluded Illinois case law did not constitute an independent and adequate state law ground requiring affirmance, however, because the Illinois Appellate Court's opinion did not contain the statement required by Michigan v. Long, 463 U.S. 1032, 1041 (1983). *Rodriguez*, 110 S. Ct. at 2798.

65. 110 S. Ct. at 2797. The prosecution made a second alternative argument

late Court affirmed the trial court's ruling in an unpublished opinion. The Illinois Supreme Court denied the prosecution's request for review.

C. *The United States Supreme Court's Decision*

The United States Supreme Court granted certiorari on both the question whether Fischer possessed authority under *Matlock* and the question whether an appearance of authority to consent could satisfy the Fourth Amendment. The Court readily decided Fischer did not meet the *Matlock* co-inhabitant standard. All of the Justices agreed

---

that, even if the police entry of Rodriguez's apartment were unconstitutional, the drugs should be admissible under the exception to the exclusionary rule created in United States v. Leon, 468 U.S. 897 (1984), because the officers "acted in reasonable, good-faith reliance on Gail Fischer's apparent authority to permit their entry." State of Illinois Brief in *Rodriguez* at 25. The suppression judge also rejected that argument, as did the Illinois Appellate Court. This issue was included in the grant of certiorari by the United States Supreme Court and was briefed, but neither the majority nor dissenting opinion in *Rodriguez* addressed this issue.

The prosecution attempted to bring the *Rodriguez* situation within *Leon*'s language by arguing the officers who entered Rodriguez's apartment relied on Fischer's apparent authority to permit their entry. While this argument mimicks the "objectively reasonable" reliance language in *Leon*, 468 U.S. at 922, it ignores the underlying ground on which *Leon* based its conclusion that the police reliance on a defective warrant was "objectively reasonable." In *Leon* the invalid search warrant was presumptively valid at the time of the search because it had been issued by a judge. *See* 468 U.S. at 921. No similar presumption arises regarding the authority of a lay third party such as Fischer.

The prosecution's argument did not simply involve an application of the existing *Leon* exception. The *Leon* exception was expressly stated in terms of evidence seized pursuant to constitutionally defective search warrants. 468 U.S. at 922 (exception applies to evidence "obtained in objectively reasonable reliance on a subsequently invalidated search warrant"). The rationale for the exception, which depends on assigning the blame for the unconsititutional search to the magistrate who issued the warrant rather than the police, was inherently limited to warrant searches. 468 U.S. at 921. The prosecution's argument in *Rodriguez* entailed not only a significant expansion of the *Leon* exception to allow it to reach a warrantless search, but also a transformation of the rationale for the exception.

It is not surprising, in view of the difficulties involved in stretching the *Leon* exception to the *Rodriguez* situation, that the Court decided *Rodriguez* on the basis of the State's "seeming consent" argument rather than its *Leon* exception argument. It is also noteworthy that the Solicitor General's Office omitted any argument in favor of extending the *Leon* exception in the brief it filed in support of Illinois' "seeming consent" argument. Amicus Brief of The United States at 14 n.6.

On the other hand, there is reason to believe at least one Justice, Justice White, the author of *Leon* and perhaps the most outspoken proponent presently on the Court of a broad good faith mistake exception to the exclusionary rule, probably would have been willing to decide *Rodriguez* by applying the *Leon* exception. Justice White took the position that the *Leon* exception should be applied to police mistakes in warrantless intrusions in his dissenting opinion in INS v. Lopez-Mendoza, 468 U.S. 1032, 1056 (1984). No other Justice has publicly espoused that view.

the State of Illinois had failed to show Fischer was a co-inhabitant.[66] Justice Scalia's majority opinion observes it is "clear" the prosecution did not show Fischer was a co-inhabitant[67] and the Illinois Appellate Court's ruling on this point was "obviously correct."[68]

The Court divided six Justices to three on the second issue of whether "seeming consent" could provide a constitutional basis for the police intrusion. Justice Scalia's majority opinion (joined by Chief Justice Rehnquist and Justices White, Blackmun, O'Connor and Kennedy) concluded a reasonable appearance of authority to consent is sufficient to satisfy the Fourth Amendment.[69] On that basis, the Court reversed the decision of the Illinois Appellate Court and remanded the case for further proceedings. Justice Marshall concluded in his dissenting opinion (joined by Justices Brennan and Stevens) that the warrantless police entry of Rodriguez's apartment violated the Fourth Amendment, regardless of the explanation for the police conduct, because there was no valid consent.

III. THE RATIONALE FOR JUSTICE SCALIA'S MAJORITY OPINION

Justice Scalia constructed the rationale for the *Rodriguez* holding from three critical doctrinal claims. His first claim is that the consent of a person whose rights are actually at stake is not an indispensible condition for an intrusion to be justified on the basis of "consent." Rodriguez's attorneys had argued Fischer's "seeming consent" had no effect on Rodriguez's right to privacy in his home under the Fourth Amendment because consent should be understood to operate as a waiver or release of a citizen's *own* privacy interests under the Fourth Amendment; therefore, Rodriguez's right could not be "vicariously waived" by Fischer.[70] Justice Scalia responded by asserting

---

66. 110 S. Ct. at 2798. This aspect of the ruling was not surprising. It was highly unlikely the Court would have granted certiorari in *Rodriguez* simply to assess the correctness of the application of the *Matlock* standard to the facts in the case.

Neither the State of Illinois nor the Solicitor General treated the *Matlock* issue prominently. The brief for the State of Illinois concentrated on the reasonable appearance-of-authority-argument and appended the *Matlock* authority-in-fact argument at the end, although the issue of Fischer's actual authority to consent under *Matlock* would seem to be the first question. The Solicitor General's amicus brief expressed the view that the Court need not address the *Matlock* issue. Amicus Brief of the United States at 20 n.16.

67. 110 S. Ct. at 2797.

68. *Id.* at 2798.

69. I refer to the *Rodriguez* standard as a "reasonable appearance" standard rather than as a "reasonable belief" standard because the Court's holding does *not* require any showing that the police actually held a subjective belief, but only that a hypothetical reasonable officer could have held one. *See supra* note 30.

70. 110 S. Ct. at 2798 (quoting Brief for Respondent at 32).

it is irrelevant that Rodriguez did not consent to the police entry himself because the Fourth Amendment does not guarantee "that no government search of [a] house will occur unless [a resident] consents; but that no such search will occur that is 'unreasonable.'"[71]

Justice Scalia's second key claim is that consent makes police intrusions constitutional by *satisfying* the Fourth Amendment's reasonableness requirement. He declares consent is one of "various elements . . . that can make a search of a person's house 'reasonable'"[72] and indicates consent searches constitute an exception to the warrant requirement that otherwise applies to searches of residences.[73] This is so, he implies, because it is "reasonable"[74] for police officers to enter a residence once they reasonably conclude they have received consent to do so.[75]

Justice Scalia's third key claim is that it does not matter if the police determination that the "consenting" person has authority to consent is erroneous provided the police error in assessing such authority is "reasonable." Justice Scalia asserts a police assessment of a third-party's authority to consent is a factual determination that should be assessed from an *ex ante* police viewpoint.[76] On that basis, he asserts that the proper inquiry is not whether the "consenting" party had authority in fact, but whether a reasonable officer could have believed the "consenting" party had authority in light of the information then available to the officer.[77] Justice Scalia then asserts that police errors about a consenting party's authority to consent fall under a "general rule"[78] that police assessments of "recurring factual questions" such as whether third parties possess authority to consent[79] need not be correct as long as they are "reasonable" ("understandable"[80] and "responsible"[81]).

Justice Scalia thus concludes a "seemingly consented search," one in which there is a reasonable appearance that the "consenting"

---

71. *Id.* at 2799.
72. *Id.*
73. *Id.* at 2797.
74. As I explain in Part V, Justice Scalia uses the term "reasonable" for Fourth Amendment analysis as though it were synonymous with "understandable in the situation." This usage is different from the traditional usage of police conduct that satisfies the probable cause requirement and the warrant requirement (or one of the enumerated exceptions to the warrant requirement). *See supra* note 21.
75. Justice Scalia's opinion in *Rodriguez* only implies this. Chief Justice Rehnquist has subsequently stated this claim more explicitly. *See infra* notes 148-50 and accompanying text.
76. *Rodriguez*, 110 S. Ct. at 2800-01.
77. *Id.* at 2799-2800.
78. *Id.* at 2800.
79. *Id.*
80. *See infra* note 254 and accompanying text.
81. *See infra* note 253 and accompanying text.

person possesses authority to consent, meets the Fourth Amendment's reasonableness requirement and is constitutional.[82] He frames the standard that reviewing courts should use to assess "seeming consent" as follows:

> determinations of consent to enter must "be judged against an objective standard: would the facts available to the officer at the moment . . . 'warrant a man of reasonable caution in the belief'" that the consenting party had authority over the premises? If not, then warrantless entry without further inquiry is unlawful unless authority actually exists. But if so, the search is valid.[83]

The rationale articulated in *Rodriguez* for upholding "seemingly consented searches" is unsupported by Supreme Court precedent,[84]

---

82. *Rodriquez*, 110 S. Ct. at 2800 n.*.

83. *Id.* at 2801 (quoting Terry v. Ohio, 392 U.S. 1, 21-22 (1968)). I address the peculiar qualification regarding actual authority and "valid" searches *infra* notes 165-71 and accompanying text. I address the weaknesses of this standard *supra* note 30.

84. There is no precedent for Justice Scalia's rationale in the decisions of the Supreme Court. It does, however, closely follow Professor LaFave's summary of the rationale offered in lower court decisions endorsing the so-called "apparent authority to consent doctrine" originating in California state law decisions:

> [It] might be said that the [apparent authority to consent doctrine] is sound in that it is coextensive with the Fourth Amendment, which protects only "against unreasonable searches and seizures." So the argument goes, if it is otherwise true that under the Fourth Amendment the police are entitled to proceed upon the basis of the "factual and practical considerations of everyday life on which reasonable and prudent men, not legal technicians, act," even if it results in a "mistake [which] was understandable and * * * a reasonable response to the situation facing them at the time," then it is likewise true that a reasonable mistake in determining a third party's authority to consent does not give rise to an unreasonable search.

LAFAVE, *supra* note 14, § 8.3(g) at 264 (footnotes omitted). Professor LaFave wrote "it is quite easy to reach the conclusion that a search should not be undone by reasonable good-faith mistakes concerning the authority of the consenting party." *Id.* at 266.

It is important, however, to note that the apparent authority to consent doctrine, which is not derived from or related to the concept of apparent authority in agency law, *see infra* note 87, arose in People v. Gorg, 291 P. 2d 469 (1955) (upholding a search in which a homeowner allowed police to enter a room in his home that was being used by a student who did gardening for the homeowner). The salient fact in *Gorg* was that the third party did have general authority over the premises; the issue involved only the scope of his authority as to a specific room. Such scope of authority cases involve different considerations than those that raise a threshold issue as to whether the third party possessed any authority at all over the premises. There is a potential for a scope of authority inquiry to degenerate into too-fine distinctions regarding the scope of a third party's authority. *See, e.g.*, Frazier v. Cupp, 394 U.S. 731 (1969), discussed *infra* note 124. The potential for such hair-splitting does not arise in a case presenting the threshold authority issue, however. *See, e.g., Rodriguez*, 110 S. Ct. 2793, 2797-98 (1990); United States v. Matlock, 415 U.S. 164 (1974). Therefore, the application of the apparent authority to consent doctrine to a police

and, I venture to say, unprincipled. Specifically, each of the three key claims from which Justice Scalia's rationale is constructed flies in the face of settled principles and doctrine.

IV. HOW *RODRIGUEZ* EVADES THE TRADITIONAL UNDERSTANDING OF CONSENT AS A FOREGOING OF A PERSON'S *OWN* PRIVACY INTEREST

Justice Scalia characterizes consent as a condition that *satisfies* the requirements of the Fourth Amendment (as does, for example, a search pursuant to a valid search warrant).[85] He asserts "consent" is an "element" amounting to a "factual determination" that makes a police intrusion "reasonable" under the Fourth Amendment.[86] His characterization of consent is significant because it focuses attention on the character of the police conduct rather than on the conduct of a person whose privacy interest is at stake. This focus-shifting move is absolutely essential to the rationale that Justice Scalia constructs in *Rodriguez*, but it is a move which demeans the nature and implications of consent.

In this part, I examine the unprecedented nature of Justice Scalia's caricature of consent by contrasting it to the traditional understanding of consent. I first set out the Court's traditional understanding of

---

error regarding a "consenting" third party who did not possess any authority over the residence is inappropriate. *Cf.*, LAFAVE, *supra* note 14, § 8.3(g) 264 ("The *Gorg* apparent authority doctrine quite clearly is not inextricably bound up with the third party consent rule set out in *Matlock* . . . .").

85. 110 S. Ct. at 2797. The Court has also recognized a variety of enumerated exceptions to the warrant requirement in which warrantless searches that meet the probable cause standard (or, in the case of minimally invasive searches, a lower reasonable cause standard) may be deemed reasonable under the Fourth Amendment. For example, the Court has long treated searches made incident to a lawful arrest as valid under the Fourth Amendment, *E.g.*, Chimel v. California, 395 U.S. 752 (1969). That doctrine was expanded in United States v. Robinson, 414 U.S. 218 (1973), to allow searches incident to custodial arrests for traffic violations. *See infra* note 169.

Another exception involves the exigent circumstances doctrine, which allows police who have probable cause to make a warrantless intrusion when there is an exigency. Exigent circumstances arise when it is infeasible to obtain a warrant because of the risk of escape of a suspect or destruction or loss of evidence. *See, e.g.*, Warden v. Hayden, 387 U.S. 294 (1967) ("hot pursuit"). Automobile searches are a specific subset of the exigent circumstances doctrine. *See, e.g.*, Chambers v. Maroney, 399 U.S. 42 (1970). Warrantless searches pursuant to either the exigent circumstances exception or the automobile exception require probable cause. *See* cases cited *infra* notes 196-97.

Another recognized exception to the warrant requirement involves limited, warrantless pat-down searches approved in Terry v. Ohio, 392 U.S. 1 (1968). *See infra* notes 205-07 and accompanying text. This listing of recognized exceptions is merely illustrative, not exhaustive.

86. *Rodriguez*, 110 S. Ct. at 2799.

consent as a citizen's waiver or foregoing of privacy and show there is no place for mere "seeming consent"[87] in the traditional concept of consent. I then show how Justice Scalia's opinion evades the traditional understanding of consent through a series of misstatements and rhetorical diversions. Finally, I demonstrate that Justice Scalia only evades the traditional concept of consent rather than repudiating it directly. I suggest this maneuver allows the majority to preserve the traditional concept of consent for application in those cases in which it would provide a better justification for upholding a police intrusion than would "seeming consent." The net result of Justice Scalia's manipulation of consent is the creation of two alternative versions of consent: "seeming consent" or consent in fact, which the Court can invoke opportunistically to uphold police intrusions.

A. The Court's Traditional Understanding of Consent as a Waiver of a Person's Own *Privacy Interest*

A number of early Supreme Court decisions treated consent instrusions as constitutional without explicitly spelling out a theory of consent.[88] Those pre-Burger Court decisions that did explain consent typically described it as a waiver.[89] The Court viewed consent to a police intrusion exactly as consent is viewed in other areas of the law—as an active form of agreement. A citizen's consent to a police intrusion was understood to waive[90] (forego or relinquish) the privacy

---

87. I use Justice Scalia's phrase "seeming consent" as a shorthand for a reasonable appearance of consent. *See supra* note 13. I avoid the term "apparent authority to consent" or "apparent consent" to avoid any suggestion that this notion is related to the concept of apparent authority in the law of agency. In agency law, apparent authority, which is a form of binding authority, can arise only from the conduct and representations of the principal, not from the conduct or representations of the agent. *See* RESTATEMENT (SECOND) OF AGENCY § 27 (1958). There could be apparent authority to consent in the agency sense only if the genuine resident (Rodriguez) had told the police that a third party (Fischer) could consent for him. The third party's own representations could never give rise to apparent authority in the agency law sense. It is unfortunate the California doctrine of appearance of authority became labeled as the apparent authority to consent doctrine. *See supra* note 84.

88. *See* Vale v. Louisiana, 399 U.S. 30, 39 (1970); Davis v. United States, 328 U.S. 582, 593-94 (1946). For a description of a number of the Court's early search cases involving consent, *see* Wefing & Miles, *supra* note 53, at 217-27.

89. Johnson v. United States, 333 U.S. 10, 13 (1948) (valid consent to police entry of residence involves "an understanding and intentional waiver of a constitutional right"); Zap v. United States, 328 U.S. 624 (1946) (defendant who consented to seizure "voluntarily waived" his Fourth Amendment right); Amos v. United States, 255 U.S. 313, 317 (1921) (the Court noted, but did not reach, the issue whether a wife could "waive" her husband's right to privacy).

90. The Court apparently stopped using the word "waiver" as a synonym for consent in Schneckloth v. Bustamonte, 412 U.S. 218, 243 n.1 (1973), to avoid any

interest of the citizen that otherwise would be protected under the Fourth Amendment.[91] Stated another way, a person's valid consent to a police intrusion provides a source of legal authority for the intrusion that would be lacking otherwise.[92]

Although the Court has said consent makes a police intrusion "constitutional," it never suggested before *Rodriguez* that consent makes a police intrusion "reasonable" under the Fourth Amendment. Rather, the Court viewed consent as taking a police intrusion outside the scope of the Fourth Amendment. The traditional notion of consent took shape before the Court's articulation in *Katz v. United States*[93] of the reasonable expectation of privacy formulation of the scope of Fourth Amendment protections.[94] The traditional notion of consent,

---

implication that a consenting person must be advised of his or her right to withhold consent. *Schneckloth* does not revise the underlying concept of consent, however; it continues to describe consent as "foregoing" a right and "allowing" an intrusion. *See infra* notes 113-20 and accompanying text.

91. *Cf.*, Lloyd L. Weinreb, *Generalities of the Fourth Amendment*, 42 U. CHI. L. REV. 47, 59 (1974) (stating with regard to a consent search that "the fourth amendment is uninterested in . . . the surrender of privacy by one who voluntarily exposes to public view what he might have kept private").

92. The justification for a police intrusion must be rooted in a source of legal authority. In the case of an intrusion into a home, the police do not usually, in the absence of exigent circumstances, possess legal authority for an intrusion unless they have obtained a warrant by showing probable cause. If the police act on the basis of a resident's consent without obtaining a warrant, the resident's consent is the sole source of legal authority for the intrusion.

93. 389 U.S. 347 (1967).

94. Although this phrase appears in Justice Harlan's concurring opinion in *Katz*, 389 U.S. at 360, it has become the standard description of the *Katz* test.

There are a substantial number of cases in which the Court has applied the *Katz* reasonable expectation of privacy standard and held the police conduct did not implicate a privacy interest protected by the Fourth Amendment. The Court's decisions in this area are complex and the rationales are varied. Two subsets, however, develop along the two prongs of the *Katz* formulation.

The first subset includes situations in which police conduct does not intrude upon any reasonable, legitimate or justifiable privacy interest. Certain interests are deemed interests that society is not willing to treat as legitimate subjects that implicate a privacy concern. For example, the Court has held no one can have a legitimate privacy interest in open fields and that the legitimate area of privacy around a house extends only to the curtilage. Oliver v. United States, 466 U.S. 170 (1984). *See also* California v. Ciralo, 476 U.S. 207 (1986) (there is no reasonable expectation of privacy in a backyard from observation from aircraft); United States v. Knotts, 460 U.S. 276 (1983) (use of an electronic tracking device does not infringe upon privacy if it reveals only information that would potentially be available to an observer).

The second subset includes situations in which the citizen has not acted in a way that demonstrates a subjective expectation of privacy. In these cases, the conduct of the citizen withdraws or defeats a privacy claim in an interest that otherwise would be a proper subject for a reasonable expectation of privacy. *Consent falls within this subset.* Consent is a citizen's agreement to expose an otherwise protectable privacy interest to the police. A citizen might also expose a legitimate subject for privacy to public view by failing to take steps to preserve privacy. *See, e.g.*, Florida v. Riley,

however, is easily subsumed under the *Katz* formulation.[95] This is so because consent amounts to a citizen's surrender of an expectation of privacy and an exposure of an otherwise private interest.[96] Under this concept, the sole focus of inquiry is whether a person whose privacy interest was at stake (a person with authority over the premises) actually gave permission for an intrusion.[97] If such a person gave consent, the Fourth Amendment's reasonableness standard is rendered inapplicable. If no such person gave consent, the protections of the Fourth Amendment apply in full.

This traditional concept of consent is readily displayed in a number of cases. It is especially evident in *Stoner v. California*,[98] *Schneckloth v. Bustamonte*,[99] and *United States v. Matlock*.[100] *Stoner* is directly

---

488 U.S. 445 (1989) (no expectation of privacy if resident exposes contents of greenhouse to observation from helicopter by leaving panels out of roof); Greenwood v. California, 486 U.S. 35 (1988) (no expectation of privacy in trash set out at curb for collection); Smith v. Maryland, 442 U.S. 735 (1979) (no expectation of privacy regarding phone numbers dialed because the citizen has made that information available to the phone company); United States v. Miller, 425 U.S. 435 (1976) (no expectation of privacy in banking records because the citizen has voluntarily conveyed that information to bank employees); United States v. White, 401 U.S. 745 (1971) (no expectation of privacy in incriminating conversation where other party is informant with radio transmitter because defendant assumed the risk that the other person could convey the information to the government); *see also* Yackle, *supra* note 2, at 355-63. Professor Yackle discusses the scope of Fourth Amendment protection and concludes the Burger Court used *Katz* to restrict the protection accorded by the amendment and employed it in a "remarkable" and "niggardly fashion." *Id.* at 362-63.

Police conduct may also fall outside of the ambit of the Fourth Amendment if it does not amount to a search or seizure. *E.g.*, California v. Hodari D., 111 S. Ct. 1547 (1991); *see infra* note 263. *See generally* Ronald Bacigal, *In Pursuit of the Elusive Fourth Amendment: The Police Search Cases*, 58 TENN. L. REV. 73 (1990). Analytically, however, the characterization of police conduct as not amounting to a seizure does not occur under the *Katz* reasonable expectation of privacy formulation.

95. Consent is an agreement to expose an otherwise protected privacy interest to the police, and that agreement ends the citizen's legitimate expectation of privacy in that interest. *See* Illinoios v. Rodriguez, 110 S. Ct. 2793, 2802 (Marshall, J., dissenting).

96. *See, e.g.*, Schneckloth v. Bustamonte, 412 U.S. 218, 243 (1972) ("there is nothing constitutionally suspect in a person's voluntarily allowing a search"). Some commentators suggest courts do not need to decide whether the Fourth Amendment applies in consent intrusions. *E.g.*, WEFING & MILES, *supra* note 53, at 215. That should be understood, however, to mean only that once a court has concluded there is consent, there is then no need for the court to inquire whether the interest would otherwise have been within the scope of Fourth Amendment protection.

97. *Cf.*, Wefing & Miles, *supra* note 53, at 215 (In assessing the validity of consent, "the issue . . . is whether the protections of the Fourth Amendment have in fact been waived by a party with authority . . . to do so.").

98. 376 U.S. 483 (1964).

99. 412 U.S. 218 (1972).

100. 415 U.S. 164 (1974).

on point with the issue raised in *Rodriguez*. In *Stoner*, police officers searched Stoner's hotel room although they had neither an arrest warrant nor a search warrant. The police obtained the help of the hotel night clerk who identified Stoner's room, opened the door with a key, and invited them to enter, saying, "Be my guest."[101] They found several articles in the room that were used as evidence in Stoner's conviction for armed robbery.[102]

The State of California argued the search of Stoner's hotel room was constitutional because it had been conducted with the consent of the hotel clerk who had either actual or apparent authority to consent to the search.[103] Justice Stewart's majority opinion first rejected the State's claim that the clerk possessed actual authority over the room.[104] Justice Stewart then also rejected the State's apparent authority claim,[105] noting:

> It is important to bear in mind that it was the petitioner's constitutional right which was at stake here, and not the night clerk's or the hotel's. It was a right, therefore, *which only the petitioner* could waive by word or deed, either directly or through an agent.[106]

---

101. 376 U.S. at 485.

102. *Id.* at 484-86.

103. *Id.* at 488. *See also* State of California Brief at 8.

104. 376 U.S. at 488. The State of California claimed the clerk had actual authority to admit the police because state law gave hotels the authority to consent to the search of a guest's room.

Justice Stewart found no such state law. *Id.* at 488, n.7. He noted such a law would not matter in any event because "[a]t least twice this Court has explicitly refused to permit an otherwise unlawful police search of a hotel room to rest upon the consent of the hotel proprietor." *Id.* at 489 (citing United States v. Jeffers, 342 U.S. 48 (1951); Lustig v. United States, 338 U.S. 74 (1949)).

105. 376 U.S. at 488-89. The State of California did not allege the clerk was Stoner's agent and its apparent authority argument does not involve the agency law concept of apparent authority. *See supra* note 87. Rather the State invoked the California apparent authority to consent doctrine, which originated in People v. Gorg, 291 P.2d 469 (1955). *See supra* note 84. The State of California argued state law permitted upholding the search "where there is evidence that the officers entertained a good-faith belief based upon sufficient facts that the person consenting to the search had authority to do so." Brief of State of California at 11, (citing, *inter alia*, *Gorg*). The California brief was conclusory in asserting that there were sufficient facts to support the alleged belief. No such facts were specifically identified.

106. 376 U.S. at 489 (emphasis added). Immediately after this passage, Justice Stewart wrote "[T]he rights protected by the Fourth Amendment are not to be eroded by strained applications of the law of agency or by unrealistic doctrines of 'apparent authority.'" *Id.* at 488. It is not clear why Justice Stewart mentioned agency law; the State had not made any agency law argument. *See supra* note 105. Perhaps Justice Stewart wanted to dispel any notion that Stoner's turning in the hotel key at the front desk whenever he left the hotel amounted to his making the clerk his agent. Justice Stewart's reference to unrealistic doctrines of apparent authority appears to be directed to the State of California's invocation of the California state law doctrine of apparent authority to consent. *See supra* note 105.

This passage, to which there was no dissent,[107] not only uses the term "waiver," but also emphasizes that only a person whose own right is implicated can possess authority to consent. It is evident that *Stoner* stands squarely against any claim that mere "seeming consent" could legitimate a police intrusion into the privacy of a home.[108]

The 1973 *Schneckloth* decision was the next major consent decision the Court rendered. *Schneckloth* addressed the question of whether voluntary consent for Fourth Amendment purposes must rest upon a "knowing and intelligent waiver" of Fourth Amendment rights.[109] In practical terms, the issue was whether the Court would require the police to warn citizens of their right to decline to consent.[110] Justice Stewart's opinion for a six Justice majority held voluntary consent could be found from the "totality of the surrounding circumstances" without proof that the consenting citizen actually knew he had a right to withhold consent.[111] The practical effect of the decision was that "the Court made it easy" for the government to establish voluntary consent.[112]

---

107. Justice Harlan's dissenting opinion in *Stoner* did not take issue with the majority opinion's treatment of the consent issue. 376 U.S. at 490-91.

108. Justice Marshall's dissenting opinion views *Stoner* as controlling authority against "seeming consent." *Rodriguez*, 110 S.Ct. at 2804-05 (Marshall, J., dissenting). Justice Scalia attempts to obfuscate the meaning of *Stoner*. *See infra* notes 162-64 and accompanying text.

109. 412 U.S. at 235. This issue arose from the Court's earlier statement that courts should "indulge every reasonable presumption against the waiver of fundamental constitutional rights." Johnson v. Zerbst, 304 U.S. 458, 464 (1938). *See also* Johnson v. United States, 333 U.S. 10, 13 (1948) (refusing to recognize consent to police entry of residence that was "granted in submission to authority rather than as an understanding and intelligent waiver of a constitutional right"). The waiver issue in *Zerbst* arose in the context of trial proceedings, not in the context of consent to a police intrusion. *Johnson*, however, applied what appears to be the strict standard of waiver to consent in a search context.

110. 412 U.S. at 245 n.33 (suggesting requiring proof that a consenting citizen was aware of his right to refuse consent would effectively require the police to give detailed warnings comparable to the *Miranda* warnings). *See also* Comment, *Consent Searches: A Reappraisal after Miranda v. Arizona*, 67 COL. L. REV. 130 (1967).

111. 412 U.S. at 235. Justice Stewart concluded "it would be next to impossible to apply to a consent search the standard of 'an intentional relinquishment or abandonment of a known right or privilege.'" *Id.* at 243.

112. Yale Kamisar, *The Warren Court (Was It Really So Defense-Minded?), The Burger Court (Is it Really So Prosecution-Oriented?), and Police Investigatory Practices, in* THE BURGER COURT: THE COUNTER-REVOLUTION THAT WASN'T 75 (Vincent Blasi, ed., 1983).

As Professor Kamisar has written:

> *Schneckloth* manifests the Court's willingness to downgrade Fourth Amendment rights . . . .
>
> Thus, when the government seeks to justify a search on "consent" grounds, it need not demonstrate a "knowing and intelligent" waiver of

Schneckloth edited "waiver" out of the Court's vocabulary of synonyms for consent to avoid any suggestion that consent required a warning of the right to withhold consent. In fact, Justice Stewart explicitly stated consent to a search did not involve a "strict standard of waiver"[113] such as would be applied to waiver of constitutional rights in a trial context.[114] Even so, *Schneckloth* did not alter the Court's basic understanding that consent can be derived only from the agreement of the person whose interest was at stake. That is evident in Justice Stewart's reference to consent as a "diluted form of waiver."[115] It is also evident in his description of consent as a "situation where a person foregoes a constitutional right"[116] and as "a person's voluntarily allowing a search."[117] As this language shows,

---

> Fourth Amendment rights — this strict standard of waiver is reserved for those rights designed to preserve a fair trial. It need only demonstrate that the consent to an otherwise impermissible search "was in fact voluntarily given, and not the result of duress, or coercion, express or implied." According to the *Schneckloth* majority, then, one may effectively consent to a search even though he was never informed—and the government has failed to demonstrate that he was aware—that he had the right to refuse the officer's "request." One need not be protected from loss by *ignorance* or *confusion*, only from loss through *coercion*. After *Schneckloth*, the criminal justice system, in some important respects at least, can (to borrow a phrase from *Escobedo*) "depend for its continued effectiveness on the citizens' abdication through unawareness of their constitutional rights."

Id. (footnotes omitted).

113. 412 U.S. at 241. Justice Stewart noted "While we have occasionally referred to a consent search as a 'waiver' we have never used that term to mean 'an intentional relinquishment or abandonment of a known right or privilege.'" *Id.* at 243 n.31. He then went on to clarify that previous cases that had used the term "waiver" analyzed the issue of consent in terms of the conduct of the consenting person and the circumstances, not in terms of the subjective understanding of the person. *Id.* That discussion clearly indicates *Schneckloth* rejects waiver only in the strict sense of the term; it does not challenge the basic proposition that consent must arise from the conduct of the person whose right is at stake.

The notion of a strict standard of waiver is also involved in Justice Stewart's statements that "a 'waiver' approach to consent searches would be thoroughly inconsistent with our decisions that have approved 'third-party consents,'" *id.* at 245, and that "it is inconceivable that the Constitution could countenance the waiver of a defendant's right to counsel by a third party." *id.* at 246. In both of these statements, Justice Stewart was using waiver in the sense of the strict standard of waiver.

114. *Id.* at 235-46. This rejection of the formal concept of waiver for Fourth Amendment consent rests on a premise that Fourth Amendment rights are of a different, and lower, order than those rights set out in other provisions of the Bill of Rights. Justice Stewart wrote "There is a vast difference between those rights that protect a fair criminal trial and the rights guaranteed under the Fourth Amendment." *Id.* at 241. *See also id.* at 242.

115. *Id.* at 245.

116. *Id.*

117. *Id.* at 243.

Schneckloth did not take issue with the basic conception that consent arises only from a citizen's giving up his or her own Fourth Amendment protections.

Significantly, no language in *Schneckloth* indicates any member of the Court viewed consent as an "element" that makes an intrusion "reasonable."[118] Nor is there any suggestion that mere "seeming consent" could substitute for consent. To the contrary, Justice Stewart expressly stated the test for valid consent is whether "the consent was *in fact* voluntarily given,"[119] not merely whether it might reasonably have appeared that way to a police officer. Justice Stewart avoided the word "waiver" simply to avoid any implication that the police had to warn the consenting party.[120]

The continuity of the Court's view of consent is also evident in the 1974 *Matlock* decision, the leading decision regarding third-party consent. Like *Schneckloth*, *Matlock* avoids the word "waiver." Even so, the analysis in the *Matlock* opinion regarding the basis for third-party consent shows the majority opinion also conceived of consent as a citizen's agreement to give up his or her *own* constitutionally protected privacy.

In *Matlock* the police arrested Matlock outside the rooming house in which he lived. They allegedly obtained the consent of his common law wife, Gayle Graff, to search the room that she and Matlock shared, and they found incriminating evidence.[121] Based on information obtained by the police primarily during and after the search,[122]

---

118. *See infra* notes 140-45 and accompanying text.

119. 412 U.S. at 248 (emphasis added).

120. *See* Wefing & Miles, *supra* note 53, at 278 ("[C]onsent constitutes a waiver of a person's constitutional rights, even accepting the position set forth in [*Schneckloth*] that it is not a waiver in the traditional [i.e., strict] sense.").

121. 415 U.S. at 177. Graff subsequently denied that she consented to the police search. *Id.* at 166.

122. Graff was present at the boarding house where Matlock lived and was clearly a resident of the boarding house (she was standing on the porch in a robe holding a baby in her arms). The key issue was whether she lived in the same room as Matlock and the evidence that linked her to that room was more limited. Justice White's opinion refers only to one piece of information the police may have obtained prior to starting their search of the room: a statement Graff allegedly made "at the time of the search" to the effect that she shared the room with the defendant. *Id.* at 166-67. It is unclear from his description, however, whether that statement was actually made before or during the search. Other information used to establish that Graff was in fact a co-inhabitant of the room appears to have been acquired *during* or *after* the search. This information included statements made by Graff after the search that she shared the room with the defendant and that they were married, the discovery by police of two pillows on the bed in the room and of both men's and women's clothing in the closet, and statements by other witnesses that Matlock and Graff cohabitated. *Id.* at 168, 169 n.3. It appears the record provided little evidence to show the police had any reasonable basis for believing, prior to the search, that Graff was a co-inhabitant of the room they wanted to search. Nevertheless, the district court ruled the officers did have a reasonable basis for such a belief when they undertook the search. *Id.* at 167.

the district court found Graff and Matlock were co-inhabitants in fact.

Two aspects of the majority opinion clearly reflect the traditional concept of consent. The first aspect is Justice White's treatment of the peculiar issue posed by third-party consent: how (or whether) the consent of one resident could affect the rights and privacy of a second resident who had not consented. Justice White did not suggest that one person could consent on behalf of another person (which would have contradicted *Stoner*'s prohibition against vicarious waiver). Neither did he deem one resident to be the agent of the other. Rather, Justice White stated Graff, who possessed common authority with Matlock as a co-inhabitant, was exercising her *own* authority when she consented to the police search. He further stated Matlock could not claim his right to privacy was violated by Graff's consent because Matlock had assumed the risk that Graff could expose the privacy of their residence to the world when he accepted Graff as a co-inhabitant.[123]

This assumption of risk language indicates the Court's view that a release of a person's right under the Fourth Amendment can be located only in an aspect of that person's *own* conduct, such as the person's acceptance of another person as a co-inhabitant.[124] Graff's

---

123. As Justice White wrote:

> The [common] authority which justifies the third-party consent . . . rests . . . on mutual use of the property by persons generally having joint access or control for most purposes, so that it is reasonable to recognize that any of the co-inhabitants has the right to permit the inspection *in his own right* and that *the others have assumed the risk* that one of their number might permit the common area to be searched.

Id. at 171 n.7 (emphasis added). This reference to a third-party's authority in his own right underscores the *Matlock* majority's insistence that the person consenting must actually possess common authority.

The Court has also employed an assumption of the risk approach in Smith v. Maryland, 442 U.S. 735, 745 (1979), United States v. Miller, 425 U.S. 435, 443 (1976) and United States v. White, 401 U.S. 745, 752 (1971).

124. Prior to *Matlock* the Court employed an assumption of risk analysis in Frazier v. Cupp, 394 U.S. 731, 740 (1969). *Frazier* involved a question of the scope of authority possessed by the co-user of a duffle bag. The Court upheld the search of a duffle bag on the consent of a joint user of the bag notwithstanding a claim that the defendant had exclusive use of the particular compartment in the bag where the incriminating evidence was found. 394 U.S. at 740. The Court declined to "engage in such metaphysical subtleties" regarding authority over the various compartments in the bag and concluded the defendant must "have assumed the risk" that the joint user of the bag had access to the entire bag.

The Court's reluctance to engage in too-fine distinctions regarding the scope of a co-user or co-inhabitant's authority is also suggested by Justice White's phrasing of the test for common authority as whether the third party "generally [has] joint access and control for most purposes." *Matlock*, 415 U.S. at 171 n.7. The *Matlock* approach does not appear to call for a detailed probing of the subjective intentions or understandings of the coresidents of a premises as to their respective areas of privacy;

consent, therefore, was effective against Matlock's right of privacy only as a result of *Matlock's own conduct* in accepting Graff as a coresident. Like *Stoner*, *Matlock* insists consent can only arise from the conduct of a person whose privacy right is actually at stake.[125]

The second aspect of *Matlock* that demonstrates the Court's continuing adherence to the traditional concept of consent is its holding. The district court had ruled a third-party consent search required the prosecution to prove both that the police had a reasonable basis to believe the third party possessed authority to consent and that the third party possessed authority to consent in fact.[126] The first prong of that showing, of course, is the "seeming consent" standard adopted in *Rodriguez*. Significantly, the *Matlock* majority dropped this reasonable appearance prong from the test for the constitutionality of a third-party consent intrusion.[127] The majority opinion held a

---

this would be an unworkable test. It seems clear the *Matlock* Court would not have upheld any claim that certain drawers in the bureau in the room shared by Matlock and Mrs. Graff were private from the co-inhabitant or outside of the co-inhabitant's authority to consent. The *Matlock* majority contents itself with a threshold determination of whether a third party has the status generally of a co-inhabitant. *Id.* at 169-71. If a third party has this status, the Court will presume the third party has general authority over the premises because the resident will be presumed to have assumed the risk of the third party's access. *Matlock* allows an expansive treatment of the scope of third-party authority once the co-inhabitant threshold is met.

125. *Cf.* Wefing & Miles, *supra* note 53, at 225 (suggesting an appearance of consent should be a legitimate basis for an intrusion only if "it is the conduct of the defendant himself that has given rise to the appearance of [third-party] authority."). This is comparable, of course, to the meaning of apparent authority in agency law. *See supra* note 87.

126. 415 U.S. at 167. The district court ruled the search of Matlock's room would only meet Fourth Amendment standards if the prosecution proved *both* "that it reasonably appeared to the searching officers 'just prior to the search,' that facts exist which [grant the third party authority to consent]" *and* that the third party possessed such authority in fact. *Id.*

The district court ruled the evidence supported a reasonable appearance to the police that the third party possessed authority, but the court found the evidence insufficient to show the third party possessed such authority in fact. *Id.* at 167-68. It reached these conclusions because it ruled the third party's statement that she co-resided with the defendant was admissible as to the first issue, but not as to the second. *Id.* The Supreme Court disagreed with this evidentiary ruling and indicated the third party's statements were admissible on the issue of her authority in fact. *Id.* at 172-77.

The district court's ruling is peculiar because the factual setting in *Matlock* is hardly propitious for a "reasonable appearance" argument; it appears the police made scant inquiries about Mrs. Graff's authority before they commenced the search of the room. The Supreme Court makes only a passing reference to a single statement made by Graff at the time of the search. *See supra* note 122. Any claim that the police behaved reasonably in the search may have evaporated when the Court clarified the co-inhabitant standard for consent.

127. Another facet of Justice White's opinion reveals the *Matlock* analysis cuts against any argument that an appearance can provide a basis for consent. Justice

showing that the "consenting party" is a co-inhabitant in fact suffices to establish valid third-party consent.[128] Thus, the *Matlock* majority rejected the view that the quality of police conduct was relevant to the assessment of the constitutionality of a consent intrusion.[129] *Matlock*'s holding is diametrically opposed to the analysis and holding in *Rodriguez*.

Stoner, *Schneckloth*, and *Matlock* all treated consent as a legal condition that could arise only from the conduct of a person whose privacy interest will be invaded.[130] None of these cases (or any other previous Supreme Court decision) supports the claim that consent is an "element" that "satisfies"[131] the standard of Fourth Amendment reasonableness.[132] None suggests mere "seeming consent" or an "appearance of consent" can hold any legal significance at all.[133]

White writes "Common authority is not to be implied from a mere property interest a third party has in the property." *Id.* at 171 n.7. *See* Chapman v. United States, 365 U.S. 610 (1961) (a landlord could not validly consent to the search of a house that he had rented to another.) The outside world might tend to regard a property interest as an indicator of authority to consent. A property interest, therefore, would appear to be an indicator of common authority to the outside world. *Matlock*, however, explicitly rejects the position that a property interest can create authority to consent; instead, it focuses on the actual assumption of risk by co-inhabitants and the actual understanding and practice of the co-inhabitants. *Matlock*'s analysis indicates there can be no consent unless there has been an assumption of risk by the co-inhabitants, regardless of any outward appearance created by a property interest.

128. *Id.* at 171. Justice White concludes the government can justify a search simply by "show[ing] that permission to search was obtained from a third party who possessed common authority over . . . the premises to be inspected." *Id.*

129. The *Matlock* Court declined to reach the government's argument that a search is reasonable if the police have a reasonable basis for believing a consenting person has authority to consent. *Id.* at 177 n.14. Nothing in the *Matlock* opinion, however, offers any support for the unaddressed government argument. *See infra* notes 151-54 and accompanying text.

130. The Court also recognized the personal nature of Fourth Amendment rights when it held that only a person whose personal privacy interests have been intruded upon has standing to invoke the exclusionary rule. *See* Rakas v. Illinois, 439 U.S. 128 (1978); Alderman v. United States, 394 U.S. 165 (1969).

131. There are occasional confused and cryptic references to consent in which the theoretical nature of consent is left ambiguous. In United States v. Katz, 389 U.S. 347 (1967), Justice Stewart wrote "A search to which an individual consents meets Fourth Amendment requirements." *Id.* at 358 n.22. This statement appears to reflect careless drafting. The sole authority Justice Stewart cites for this statement is Zap v. United States, 328 U.S. 624 (1946). *Zap*, however, upheld a warrantless seizure on the grounds that "petitioner . . . voluntarily waived such claim to privacy which he otherwise might have had . . . ." *Id.* at 628.

132. As I explain in Part IV, consent cannot satisfy the Fourth Amendment because consent does not satisfy either the probable cause or warrant requirements.

133. Professor Wienreb previously concluded that a mere appearance of authority could not suffice for consent: "The answer to the question of whether a searching officer's reasonable belief that a consenting party had authority is sufficient to uphold a consent search is clearly no. 'Apparent authority' to consent is not by

Under the traditional understanding of consent, therefore, the police entry of Rodriguez's apartment was clearly nonconsenual. Rodriguez, the only inhabitant of the apartment, had not given consent, and Fischer's "seeming consent" could not provide any legal authority for the intrusion. As Justice Marshall wrote in his *Rodriguez* dissent, "Even if the officers reasonably believed that Fischer had authority to consent, she did not, and Rodriguez's expectation of privacy was therefore undiminished."[134]

B. Rodriguez*'s New Version of "Consent"*

Justice Scalia responds to the obstacle posed by the traditional concept of consent by inventing a new, more malleable version of consent. He does not acknowledge this invention; he merely insinuates it as though it were the way the Court has always viewed consent.

1. *Rodriguez*'s Claim that Consent is an "Element" That Makes an Intrusion "Reasonable"

Justice Scalia revises the Court's understanding of consent by treating it as an "element" that *satisfies* the requirements of the Fourth Amendment. Asserting this critical claim in a rather low-key manner, Justice Scalia writes: "There are various elements, of course, that can make a search of a person's house 'reasonable'—one of which is the consent of the person or his cotenant."[135] This treatment of consent provides Justice Scalia with a platform for justifying "seeming consent." Justice Scalia proceeds to assert that, as an "element," consent is to be assessed from the *ex ante* viewpoint of the police officer.[136] The adoption of that viewpoint, in turn, opens

---

itself a basis for sustaining a search . . . ." Weinreb, *supra*, note 91, at 64. Professor White reached the same conclusion:

> How should the Court deal with a case in which the consenting third party does not in fact . . . have any . . . actual authority to consent, but in which the police reasonably believe him to . . . have such authority? Since the analysis I propose is not a way of assessing the reasonableness of police conduct but a way of defining legitimate expectations of privacy, a search pursuant to such consent could not be valid unless it is the conduct of the defendant himself that has given rise to the appearance of authority.

White, *supra* note 53, at 225.

134. *Rodriguez*, 110 S. Ct. at 2804 (Marshall, J., dissenting).

135. *Id.* at 2799. No authority is cited for this statement.

136. The adoption of the *ex ante* police viewpoint is evident in Justice Scalia's statement that:

> The only basis for contending that the constitutional standard could not possibly have been met [in the police entry of Rodriguez's apartment on the basis of Fischer's "seeming consent"] is the argument that reasonableness

the way for his treatment of "seeming consent" as the equivalent of "consent." The unprecedented nature of Justice Scalia's version of consent is evident from his inability to provide authority for it. He offers few citations, and those he does offer do not support his claims.

Justice Scalia lays the foundation for his claim that consent is an "element" that satisfies the Fourth Amendment by asserting that consent operates as a recognized exception to the warrant requirement. He begins this discussion by correctly noting "[t]he Fourth Amendment generally prohibits the warrantless entry of a person's home . . . ."[137] He then adds the following:

> The prohibition [against the warrantless entry of a person's home] does not apply, however, to situations in which voluntary consent has been obtained, either from the individual whose property is searched, see *Schneckloth v. Bustamonte*, 412 U.S. 218 (1973), or from a third party who possesses common authority over the premises, see *United States v. Matlock*, 415 U.S., at 171.[138]

This language may leave the reader with the impression that *Schneckloth* and *Matlock* treated consent as one of the recognized exceptions to the warrant requirement.[139] This impression in turn may suggest they assessed consent in terms of Fourth Amendment reasonableness,[140] but neither of the cited cases treated consent that way.

Justice Scalia does not cite a specific page in *Schneckloth*, but this is the most relevant passage:

> It is well settled . . . that a search conducted without a warrant issued upon probable cause is "*per se* unreasonable . . . subject only to a few specifically established and well-delineated exceptions." It is equally well settled that one of the specifically established exceptions to the requirements of both a warrant and probable cause is a search that is conducted pursuant to consent. . . . *Zap v. United States*, 328 U.S. 624, 630.[141]

---

must be judged by the facts as they were, rather than by the facts as they were known [by the officers at the moment].

110 S. Ct. at 2800 n.*. Justice Scalia also incorporates an *ex ante* police viewpoint in his statement of the objective standard to be applied to assess "seeming consent": "would the facts available to the officer at the moment . . . ." *Id.* at 2801.

137. *Id.* at 2797 (citing Payton v. New York, 445 U.S. 573 (1980); Johnson v. United States, 333 U.S. 10 (1948)).

138. 110 S. Ct. at 2797.

139. The warrant requirement is discussed *infra* text accompanying notes 189-95.

140. The implication in this passage is a key aspect of the *Rodriguez* rationale. Note, however, that Justice Scalia only implies this claim.

141. *Schneckloth*, 412 U.S. at 219 (citations omitted). *See infra* notes 149-50 and accompanying text.

This statement does not mean consent is an exception to the warrant requirement in the sense that an exigent circumstance constitutes a recognized exception to the warrant requirement.[142] Rather, the statement that consent is an exception to *both* the warrant *and* the probable cause requirements means the Fourth Amendment reasonableness standard does not apply to consent searches. That is the meaning of the statement because "the requirements of both a warrant and probable cause" *constitute* Fourth Amendment reasonableness.[143]

Justice Stewart apparently adopted this indirect language to avoid the term "waiver." Unfortunately, his language has produced some confusion.[144] Nevertheless, *Schneckloth*'s citation of *Zap* leaves no doubt about the meaning of the passage because *Zap* expressly described the petitioner as having consented to a seizure because "he voluntarily waived such claim to privacy which he otherwise might have had . . . ."[145]

The *Matlock* opinion also does not say consent searches are an exception to the warrant requirement. The statement Justice Scalia cites says only that "'third party consent' searches" previously have been recognized to be "constitutionally valid."[146] As the cited passage does not offer any explanation why that is so, there is no reason to read that particular statement as endorsing any novel version of

---

142. An exception to the warrant requirement such as exigent circumstances would still require probable cause for entry of a home. *See infra* notes 224-25 and accompanying text.

143. The core of the reasonableness standard is the cause requirement. If the probable cause requirement is inapplicable, then the reasonableness standard of the Fourth Amendment is inapplicable. *See infra* Part IV.

144. A number of commentators have misread *Schneckloth* as though it treated consent searches as reasonable. For example, Professor Coombs has suggested the *Schneckloth* Court "determined that the consent made the search reasonable and therefore valid." Professor Coombs bases this statement on the passage from *Schneckloth* quoted *supra* in the text accompanying note 144. Mary I. Coombs, *Shared Privacy and the Fourth Amendment, or the Rights of Relationships*, 75 CALIF. L. REV. 1593, 1640 (1987). Professor Goldberger has misread *Schenckloth* in the same way. He wrote that "the Court found a search reasonable based on a third-party's voluntary though uninformed consent." He does not offer any citation to *Schneckloth* (or any other case) as authority for that statement. *See* Goldberger, *Consent, Expectation of Privacy and the Meaning of "Searches" in the Fourth Amendment*, 75 J. CRIM. L. & CRIMINOLOGY 319, 328 (1984). Professor Goldberger repeatedly assesses consent searches in terms of "reasonableness," *e.g.*, *id.* at 356, 359. Student commentators have also misread *Schneckloth*. One writes that *Schneckloth* stands for the proposition that consent is a well settled exception to the warrant requirement, but later refers to "the third-party consent search exception to the Fourth Amendment." The two descriptions are not equivalent: the latter is correct, but the former is misleading insofar as it implies that consent has been analyzed under the reasonableness standard. Comment, *Third-Party Consent Searches, The Supreme Court, and the Fourth Amendment*, 75 J. CRIM. L. & CRIMINOLOGY 963, 963-64, 991 (1984).

145. *Zap*, 328 U.S. at 628.

146. *Matlock*, 415 U.S. at 171.

consent. As explained above, *Matlock*'s holding and "assumption of risk" analysis are clearly derived from the traditional understanding of consent.[147] In sum, the only two authorities Justice Scalia cites do not stand for the proposition he maintains (or at least implies). Never before has the Court characterized consent as an exception to the warrant requirement.

It is extremely significant, moreover, that Justice Scalia does not offer any authority at all for his crucial claim that consent makes an intrusion "reasonable."[148] He did not cite authority for that notion because there is none. In fact, Chief Justice Rehnquist indirectly demonstrated the absence of authority for this point in his subsequent opinion in *Jimeno*. The Chief Justice cited only *Schneckloth* as authority for his statement that "[W]e have long approved consensual searches because it is no doubt reasonable for the police to conduct a search once they have been permitted to do so."[149] *Schneckloth* clearly does stand for the proposition that the Court has "long approved consensual searches," but the rest of the passage is entirely the invention of Chief Justice Rehnquist. His citation to *Schneckloth* is to the same passage quoted above, but that passage does not say that consent makes an intrusion "reasonable." To the contrary, *Schneckloth* treated consent as *an exception to the reasonableness standard* which hardly carries the meaning Chief Justice Rehnquist ascribed to it.[150]

Justice Scalia also attempts to give plausibility to his treatment of "seeming consent" in *Rodriguez* by framing the issue as one that "we expressly reserved in *Matlock*."[151] He refers to a footnote in *Matlock* which states:

> [W]e do not reach another major contention of the United States in bringing this case here: that the Government in any event had only to satisfy the District Court that the searching officers reasonably believed that [the third party] had sufficient authority over the premises to consent to the search.[152]

---

147. *See supra* text accompanying notes 121-28.

148. *See supra* text accompanying notes 137-43.

149. Florida v. Jimeno, 111 S. Ct. 1801, 1803 (1991) (citing Schneckloth v. Bustamonte, 412 U.S. 218, 219 (1973)). The *Jimeno* majority adopts an expansive approach to the scope of consent—consent to a search of a car amounts to consent to search of the contents of containers found in the car. *Id.*

150. *See supra* quoted passage in text accompanying note 141 and text accompanying notes 142-45. *Id.*

151. *Rodriguez*, 110 S. Ct. at 2796.

152. 415 U.S. 164, 177 n.14. The *Matlock* footnote refers to a situation in which the searching officers reasonably believed the consenting party had authority. The language of the objective standard adopted in *Rodriguez*, however, does not require any actual belief on the part of the police regarding the authority of the third party. *See supra* note 30.

This footnote, however, merely acknowledges an unaddressed argument; it does not offer any basis for thinking that the government's "reasonable belief" argument could suffice for a valid consent search,[153] especially in view of the fact that *Matlock* analyzes authority to consent solely in terms of actual authority and assumption of risk.[154]

Thus, there is not an ounce of authority for Justice Scalia's claim that consent is an "element" that makes an intrusion "reasonable."[155] There is no authority for the foundational claim on which the entire *Rodriguez* rationale depends.

2. How the Majority Opinion Begs the Crucial Issues Regarding Consent

The unprecedented nature of Justice Scalia's claims about the nature of consent is also reflected in the diversionary character of his attack on the arguments made in Justice Marshall's dissent. For example, Justice Marshall's dissenting opinion correctly points out that consent intrusions have not been assessed in terms of their "reasonableness" because consent has been understood to be a condition that makes the Fourth Amendment inapplicable.[156] Rather than respond directly to this point, Justice Scalia characterizes Justice Marshall's position as a claim that a "[consented] search becomes not really a search at all," and he then mocks that claim by writing "[t]o describe a consented search as a noninvasion of privacy and thus a nonsearch is strange in the extreme."[157]

Justice Scalia misstates Justice Marshall's analysis. Justice Marshall does not argue a consented intrusion does not amount to a search or seizure in terms of the nature of the police conduct; he argues consent eliminates the consenter's "reasonable expectation of privacy."[158] There is nothing at all "strange" about the analysis that Justice Marshall actually makes.[159]

---

153. This footnote may reflect the interest of its author, Justice White, in the possibility of excusing police conduct on the basis of reasonable mistakes. Shortly after *Matlock* was decided, Justice White emerged as one of the most outspoken advocates on the Court for a broad good-faith mistake exception to the exclusionary rule. *See supra* note 65.

154. *See Matlock*, 415 U.S. at 169-72. The holding in *Matlock* rejects the idea that the character of police conduct has any relevance for the constitutional validity of a consent intrusion. *Id.*; *see supra* notes 126-29 and accompanying text.

155. The Court has consistently treated valid consent as a condition that places the police intrusion outside the scope of the Fourth Amendment's protections. *Rodriguez*, 111 S. Ct. at 2804 (Marshall, J., dissenting).

156. *Id.* Regarding the lack of precedent for the majority approach, Justice Marshall stated in his dissent "Our prior cases discussing searches based on third-party consent have never suggested that such searches are 'reasonable.'" *Id.*

157. 110 S. Ct. at 2800 n.*.

158. *Id.* at 2802 (citing Katz v. United States, 389 U.S. 347, 351 (1967)). Justice Marshall's citation to *Katz* makes the content of his argument unmistakeable.

159. Justice Scalia's "strange in the extreme" characterization would be better addressed to his own conclusion in California v. Hodari D., 111 S. Ct. 1547 (1991), that a police chase of a suspect does not amount to a seizure. *See infra* note 263.

Justice Scalia's response to the defendant's reliance on *Stoner* is similarly diversionary. Rodriguez's attorneys argued *Stoner* precludes any claim that Fourth Amendment rights can be "vicariously waived."[160] Justice Scalia responded by simply reiterating *Schneckloth*'s rejection of the strict standard of waiver.[161] That response begs the issue. The defendant did not invoke "waiver" as a magic word, or argue that consent is invalid in the absence of a warning. The defendant's point was that consent must be based on the conduct of the person whose privacy is at stake. Justice Scalia does not address that point directly.

Justice Scalia does claim a later passage in the *Stoner* opinion made *Stoner*'s view of consent "ambiguous . . . perhaps deliberately so."[162] This, however, manufactures "ambiguity" from a simple example of judicial overkill.[163] The obvious reading of *Stoner* is the correct reading; there is no ambiguity in *Stoner*'s insistence that

---

160. 110 S. Ct. at 2798 (quoting Respondent's Brief at 32). Respondent made this argument on the basis of the use of the term "waiver" in the passage in *Stoner* quoted *supra* in the text accompanying note 106.

161. 110 S. Ct. at 2798-99. Justice Scalia reiterates Justice Stewart's discussion in *Schneckloth* regarding the distinction between the strict standard for waiver of trial rights and the lower standard applicable to voluntary consent to a search.

162. *Id.* at 2801. Justice Scalia bases his claim of ambiguity on the following passage from *Stoner*:

> It is true that the night clerk clearly and unambiguously consented to the search. But there is nothing in the record to indicate that *the police had any basis whatsoever to believe that* the night clerk had been authorized by the petitioner to permit the police to search the petitioner's room.

110 S. Ct. at 2801 (quoting Stoner v. California, 376 U.S. at 489 (1964)) (emphasis added by Justice Scalia).

Justice Scalia argues that the italicized language shows that *Stoner* does not rule out the possibility of upholding a third-party consent search on the basis of an appearance of authority because these words "should have been deleted [by Justice Stewart], of course, if the statement two sentences earlier meant that an appearance of authority could never validate a search." *Id.*

Justice Scalia's error is in failing to appreciate that Justice Stewart's language may simply reflect his concern with the implications of the agency law issue of apparent authority, *see supra* note 106, rather than a mere appearance of authority as in *Rodriguez*. If there were facts in the record that showed Stoner had in some way manifested to the police that the clerk was his agent, then the agent would have possessed valid apparent authority as an agent. Stewart, in effect, notes that there is "nothing in the record" to support any such valid agency. This has no relevance to *Rodriguez*, however, because there is no issue of agency authority in *Rodriguez*. To the contrary, *Rodriguez* only involves a claim of an appearance that Fischer had authority in her own right.

In addition, it appears that Justice Stewart was inclined to include unnecessary statements in his writing. For example, there is another example of this sort of overkill in Justice Stewart's treatment of the actual authority issue in *Stoner*. He first wrote that there was no state law conferring authority on hotel personnel and then wrote that, even if there were such a law, it would be unconstitutional in any event. *See supra* note 104. The first of these statements is unnecessary in light of the second and also could have been deleted.

163. *See* 110 S. Ct. at 2805 n.1 (Marshall, J., dissenting).

consent must come from the person whose right is at stake.[164]

C. Rodriguez's *Reservation of the Traditional Understanding of Consent*

The rationale for "seeming consent" in *Rodriguez* rests on premises that are inconsistent with the traditional concept of consent. Even so, *Rodriguez* does not actually repudiate that traditional understanding of consent. The reason for this is evident in the *Rodriguez* holding: the majority wants to keep the traditional conception available for those cases in which it would be uniquely useful for upholding consented searches.

This motive is evident in the cryptic qualification that Justice Scalia adds at the end of his statement of the standard that is to be applied in assessing "seeming consent." He writes that if the facts available to the officer at the moment would not warrant a reasonable belief that the third party possessed authority to consent (that is, if there were no "seeming consent"), "then warrantless entry without further inquiry is unlawful *unless authority actually exists. But if so, the search is valid.*"[165] In other words, after formulating a rationale that effectively repudiates the traditional conception of consent, Justice Scalia sneaks it back in so it still may be used in those consent cases in which an intrusion cannot be defended as "reasonable" under "seeming consent"!

The pragmatic reason for the insertion of this unexplained qualification is apparent; it is needed to avoid overruling prior cases such as *Matlock* in which the Court applied the traditional concept of consent.[166] This qualification clearly will be useful for upholding some consent intrusions that could not pass a "seeming consent" standard. Undoubtedly there are cases in which the police obtain valid consent from a third party who does have authority to consent but in which the police fail to obtain any information to warrant a reasonable belief to that effect before the intrusion.[167]

---

164. *See supra* text accomanying note 106.

165. *Rodriguez*, 110 S. Ct. at 2801. *See supra* text accompanying note 83.

166. *Matlock* expressly held a consent intrusion was constitutional if consent was given by a person who possessed authority to give it and that it was irrelevant whether or not the police obtained information to that effect prior to receiving the consent or undertaking the intrusion. *See supra* notes 126-29 and accompanying text. If *Rodriguez* had not included this qualification, it would have totally overruled *Matlock*.

167. The facts in *Matlock* approach this scenario. *See supra* note 122. It is not difficult to imagine situations in which the valid consent standard in *Matlock* would be more useful for upholding a third-party consent intrusion than the "seeming consent" standard in *Rodriguez*.

Assume police officers go to a house where a noisy party is in progress and

Justice Scalia does not offer a word of explanation regarding this qualification. Rather, he merely applies a semantic fig leaf. He does not write that actual authority to consent makes a search "reasonable"; he writes that it makes a search "valid." This ploy is not enough to hide the naked inconsistency between this qualification and Justice Scalia's claim that consent must be evaluated from an *ex ante* police viewpoint.[168] In fact, this qualification is blatantly inconsistent with his claim that consent operates as a form of Fourth Amendment reasonableness.

How can consent be "valid" under Justice Scalia's formulation of consent as an element that makes an intrusion "reasonable" if the police assessment of authority to consent cannot be said to be "reasonable"[169] (even under a loose interpretation of that term)? The

---

knock on the door. They say to the person who answers "We'd like to talk to whoever is in charge." The person who answered the door says, "Come on in, he's in back." The police enter. This police entry does not appear justifiable under *Rodriguez* because the police did not have any information regarding the status of the third party who "consented." In fact, the response of the person who answered the door suggests that person probably did not have authority. If it turns out the person who answered the door actually was a co-inhabitant, however, the police entry could be upheld under the actual authority standard in *Matlock*.

168. If the inquiry in consent intrusions is whether the police acted "reasonably" (understandably), it should be immaterial whether there is genuine authority to consent or not. Only "seeming consent" is relevant to assessing the "reasonableness" (understandableness) of the police conduct. The valid search qualification can only be derived from the traditional understanding of consent as a waiver. *See* LAFAVE *supra* note 14, at § 8.3(g) at 269.

169. Justice Scalia does not mention one possibity: then-Justice Rehnquist's majority opinion in United States v. Robinson, 414 U.S. 218 (1973), labeled a search of the contents of a cigarette wrapper as a "reasonable" search incident to an arrest for a traffic violation even though there was no cause for the search. The search could not have produced additional evidence of the traffic violation, and the officer did not fear the defendant was armed (the cigarette package could not have held a weapon in any event). Earlier searches incident to arrest had been justified on the grounds that such searches were needed to prevent the destruction of evidence or to protect the officer. *E.g.*, Chimel v. California, 395 U.S. 752 (1969). Justice Rehnquist did not attempt to argue there was any form of cause for the search; he argued instead the search was reasonable because it was made pursuant to authority arising from the officer's taking the driver into custody. He stated "it is the fact of custodial arrest which gives rise to the authority to search . . . ." 414 U.S. at 236.

One could analogize from *Robinson* that if authority, even standing alone, can make a search reasonable, then valid authority to consent, even standing alone, could make a consented search reasonable. All this really shows, however, is that Justice Rehnquist's rationale for *Robinson* is strange. Authority alone does not constitute reasonableness. If authority alone can make a search constitutional, regardless of an evaluation of cause for the search, then what is really being asserted is that the arrested person has lost any reasonable expectation of privacy under the *Katz* formulation. *See supra* notes 93-96 and accompanying text; *see also* Yackle, *supra* note 2, at 401.

Robinson would have done less damage to constitutional doctrine (though the

Rodriguez majority's willingness to employ an either/or formulation of two logically inconsistent treatments of consent[170] in a single opinion, indeed, in a single holding, is a disturbing illustration of the current majority's zeal for multiplying the options for upholding police intrusions.[171]

In summary, the net result of the manipulation of "consent" in *Rodriguez* is the provision of two alternative, but logically contradictory, theories that courts can opportunistically employ to uphold "consented" police intrusions into citizens' homes. The only principle evident in *Rodriguez*'s treatment of "consent" is heads-the-government-wins/tails-the-defendant-loses.[172] *Rodriguez*'s treatment of "con-

---

immediate implications for the driver's privacy would be the same) if it had simply declared that society is not willing to recognize that an arrested person has any legitimate privacy interest in their person. Justice Powell, describing what he terms "the essential premise of our decisions," wrote: "The search incident to arrest is reasonable under the Fourth Amendment because the privacy interest protected by [the Fourth Amendment] is legitimately abated by the fact of the arrest." 414 at 237-38.

This statement is also flawed. If the privacy interest is abated, the search incident to arrest falls outside of the scope of the amendment as defined by *Katz*, and there is neither a need nor a basis for labeling the search reasonable. If the word "reasonable" in Justice Powell's statement is changed to "constitutional," it would then be a statement of a *Katz* analysis.

Of course, there is an important distinction between the analysis in *Robinson* and that which would be required for the situation in *Rodriguez*; namely, *Robinson* addresses a situation involving *valid* legal authority for the arrest. Hence, the *Robinson* approach could not legitimate an intrusion based on mere "seeming" legal authority to consent.

170. 110 S. Ct. at 2801. They are logically inconsistent because the premises that are needed for one are the opposite of the premises needed for the other.

171. While it is unusual to find this lack of principle so evident within a single opinion, other instances of result-driven contradictions that yield consistently pro-government outcomes can be identified in recent Fourth Amendment decisions. For example, the Burger and Rehnquist Courts have exploited inconsistent treatments of the personal nature of Fourth Amendment rights in decisions affecting the application of the exclusionary rule. On the one hand, the Court has ruled there is no personal right to have unconstitutionally seized evidence excluded because the exclusionary rule is only a social policy aimed at deterring future unconstitutional police conduct. United States v. Calandra, 414 U.S. 338 (1974). On the other hand, when it comes to defining who can invoke the exclusionary rule, the so-called standing requirement, the Court has ruled only a person whose personal privacy interests have been violated may seek exclusion of unconstitutionally seized evidence because Fourth Amendment rights are personal, even though elimination of the standing rule probably would increase the deterrent efficacy of the exclusionary rule. *See* United States v. Alderman, 394 U.S. 165 (1969). *See generally* Burkoff, *supra* note 2.

172. A similar display of have-it-both-ways inconsistency may be involved in the position Chief Justice Rehnquist and Justice Blackmun took in Minnesota v. Olson, 110 S. Ct. 1684 (1990), which was decided after the oral argument was held but before the decision was announced in *Rodriguez*. The key issue in *Olson* was whether an overnight guest has a sufficient "legitimate expectation of privacy" in the premises to have standing to challenge the constitutionality of a police entry of

sent" does not rest on logic, doctrine, precedent, or principle; it is entirely result driven. Its implications, however, are far reaching. By deriving "consent" from the conduct of a nonresident, *Rodriguez* dilutes citizens' control over access to their homes. This is an abrupt departure from the traditional understanding that the Fourth Amendment conveys a right to citizens "to be secure in their . . . houses."

V. HOW *RODRIGUEZ* DOWNGRADES THE FUNDAMENTAL CONCEPT OF FOURTH AMENDMENT REASONABLENESS

On the surface, *Rodriguez* is about the nature of consent. Just below the surface, however, *Rodriguez*'s treatment of consent implicates the nature of the reasonableness standard contained in the Fourth Amendment. In this part, I examine *Rodriguez*'s treatment of the concept of Fourth Amendment reasonableness. Justice Scalia's majority opinion conveys the impression that the decision is based on the Court's traditional understanding of reasonableness. It is not.

The traditional understanding of Fourth Amendment reasonableness requires that an intrusion of a home, which is inherently highly invasive of a citizen's privacy, must be based on probable cause and must be authorized by a magistrate. *Rodriguez*, however, labels an intrusion based only on "seeming consent" as "reasonable" even though such an intrusion meets neither the probable cause requirement nor the warrant requirement. Justice Scalia's claim that a "seemingly

---

the premises. *Id.* at 1687. Language suggesting that guests would have standing had appeared in McDonald v. United States, 335 U.S. 451 (1948). *See also* Alderman v. United States, 394 U.S. 165, 173 n.7 (1969). However, then-Justice Rehnquist's opinion in Rakas v. Illinois, 439 U.S. 128, 142 (1978), casted doubt on the status of a guest for standing purposes when it suggested a guest would not have standing simply because the guest was "legitimately on the premises."

The Court's consideration of *Olson* and *Rodriguez* at the same time created a potential tension between the Court's treatment of the overnight guest's standing in *Olson* and of an occasional guest's "seeming consent" in *Rodriguez*. The Court would have been blatantly inconsistent had it ruled an overnight guest did not have a sufficient relationship to a residence to have standing in *Olson* and then ruled a few weeks later that an occasional guest's "seeming consent" could make a police intrusion constitutional in *Rodriguez*.

As it turned out, the Court ruled seven to two in *Olson* that an overnight guest does have standing. 110 S. Ct. at 1689. The majority decided the threshold for standing is less demanding than the *Matlock* co-inhabitant threshold for authority to consent to a search. *Id.* at 1690. The outcome in *Olson* is interesting because Chief Justice Rehnquist and Justice Blackmun, who both voted in the majority in *Rodriguez*, dissented in *Olson*. They did not explain their votes, however; neither wrote an opinion. *Id.* at 1690. It is thus possible they dissented because they objected to allowing an overnight guest to have standing. This appears to be a likely explanation because Justice Blackmun previously joined Justice Rehnquist's opinion in Rakas v. Illinois, 439 U.S. 128 (1978). It is possible these two Justices adopted inconsistent approaches to *Olson* and *Rodriguez*.

consented search" of a home is "reasonable" is both arbitrary and radical. Its colloquialized treatment of "reasonableness" carries detrimental implications for citizens' privacy and trivializes the meaning of the Fourth Amendment.

A. Fourth Amendment Reasonableness

The significance of *Rodriguez*'s claims regarding the "reasonableness" of an intrusion based on "seeming consent" is best understood in the context of the meaning of the phrase "unreasonable searches and seizures" in the Fourth Amendment. The Fourth Amendment provides:

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.[173]

The "syntactical mystery"[174] of the relationship between the first and second clauses in the amendment has been the subject of debate. In particular, controversy has centered on whether or how the specific requirements for the issuance of warrants in the second ("warrant"[175]) clause inform the meaning of reasonableness in the first clause.[176] The

---

173. U.S. CONST. amend. IV.

174. H. Richard Uviller, *Reasonability and the Fourth Amendment: A (Belated) Farewell to Justice Potter Stewart*, 25 CRIM. L. BULL. 29, 33 (1989).

175. The "warrant clause" label is unfortunate because it implies the standards set out in that clause pertain only to warrant intrusions. *See infra* notes 198-99. This label is especially unfortunate in view of the actual infrequency of warrant use. Relatively few police intrusions are conducted pursuant to warrants. The large proportion of police intrusions are justified in terms of consent or one of the recognized exceptions to the warrant requirement that the Court has developed. *See supra* note 17. Accordingly, labeling the second clause the warrant clause gives it an appearance of limited relevance. Properly understood, the cause requirement set out in the second clause is the central substantive requirement that applies even to warrantless police intrusions. *See infra* notes 233-36 and accompanying text.

176. *See* Robert M. Bloom, *The Supreme Court and Its Purported Preference for Search Warrants*, 50 TENN. L. REV. 231 (1983); Joseph D. Grano, *Rethinking the Fourth Amendment Warrant Requirement*, 19 AM. CRIM. L. REV. 603 (1982); Yale Kamisar, *Does (Did) (Should) the Exclusionary Rule Rest on a "Principled Basis" Rather Than an "Empirical Proposition"?*, 16 CREIGHTON L. REV. 565, 571-79 (1982-83); Silas J. Wasserstrom, *The Fourth Amendment's Two Clauses*, 26 AM. CRIM. L. REV. 1389 (1989); White, *supra* note 53, at 172-73; Yackle, *supra* note 2, at 385-89. For additional discussion of three possible readings of the two clauses of the amendment, see JACOB LANDYNSKI, SEARCH AND SEIZURE AND THE SUPREME COURT: A STUDY IN CONSTITUTIONAL INTERPRETATION 42-43 (1966).

"legislative history" regarding the amendment's wording is too skimpy to answer the question.[177]

Two ways of reading the two clauses in the amendment have emerged.[178] One interpretation, which I call the "traditional" reading of the amendment because it has been endorsed by the Court for a substantial period of time, views the second clause as carrying important implications for the meaning of the reasonableness requirement. This is not to say the traditional reading endorses a rigid warrant requirement. To the contrary, there are textual,[179] historical,[180] and practical reasons[181] not to interpret the amendment as always requiring a warrant for a government search or seizure.[182] The traditional reading, however, does treat warrant searches as the model for reasonable searches.

The other interpretation, which I call the "generalized reasonableness" reading, would treat the reasonableness requirement in the first clause as a free-standing standard that is not mediated by the second clause. This second interpretation is a persistent dissenting strain in the Court's opinions that is clearly becoming more visible

---

177. For brief discussions of the debate concering the wording of the Fourth Amendment, see Martin Grayson, *The Warrant Clause in Historical Context*, 14 AM. J. CRIM. L. 107, 116-17 (1987); Silas J. Wasserstrom & Louis M. Seidman, *The Fourth Amendment as Constitutional Theory*, 77 GEO. L. J. 19, 79-84 (1988); Yackle, *supra* note 2, at 337-44.

178. *See, e.g.*, Uviller, *supra* note 174, at 33-36. Professor Uviller labels the two interpretations as the "categorical" approach (which I call the traditional interpretation) and the "pure reasonability" approach (which I call generalized reasonableness). *See also* Craig Bradley, *Two Models of the Fourth Amendment*, 83 MICH. L. REV. 1468 (1985). Professor Bradley discusses the "bright lines" model of the Fourth Amendment (roughly comparable to the traditional interpretation) and the "no lines" model (roughly comparable to generalized reasonableness). *See also* Yackle, *supra* note 2, at 389-90.

179. If the Framers did not intend an intrusion could ever be reasonable without a warrant, it is difficult to explain why the Framers bothered to include a general statement of the reasonableness standard in the first clause.

180. A rigid warrant requirement appears to be implausible in the face of history because certain government officials were empowered to undertake specific types of intrusions without obtaining warrants at the time of the framing of the Bill of Rights. TELFORD TAYLOR, TWO STUDIES OF CONSTITUTIONAL INTERPRETATION, 27-29 (1969) (warrantless arrests of suspected felons, as well as searches of the person of the arrestee, were recognized at common law); Gerard V. Bradley, *Present at the Creation?, A Critical Guide to Weeks v. United States and its Progeny*, 30 ST. LOUIS U.L. REV. 103, 104 n.64 (1986) (listing a variety of searches and seizures that were allowed without warrant at the time the amendment was drafted).

181. There are times when it clearly is infeasible for police to obtain a magistrate's approval before taking necessary actions to protect public safety. This practical concern is reflected in the exigent circumstances doctrine. *See* case cited *infra* note 197.

182. As a result, there are few advocates of a rigid warrant requirement. *But see* Grayson, *supra* note 177.

in the Rehnquist Court.[183] I argue below that although *Rodriguez* is written as though it were decided under the traditional understanding of the amendment, it is actually an expression of the generalized reasonableness reading.

1. The Court's Traditional Understanding of Fourth Amendment Reasonableness

The traditional interpretation of the Fourth Amendment derives from the premise that the Framers intended warrant intrusions to be the model for reasonable searches and seizures.[184] In other words, the traditional reading of the amendment takes the view that the second clause of the amendment should be understood as a more specific and concrete expression of what the Framers intended the reasonableness requirement in the first clause should usually (though not rigidly) mean.[185]

The Court's treatment of a warrant intrusion as the model for a reasonable intrusion carries two important implications. The first is that warrantless intrusions should not be encouraged; rather, the opportunities for warrantless intrusions should be confined carefully and limited to narrowly defined circumstances. The second implication of treating warrant intrusions as the model for reasonable intrusions is that warrantless intrusions should be assessed, insofar as possible, according to the same standards that are explicitly mandated for the issuance of warrants. For that reason, the Court has recognized the probable cause standard as the benchmark for assessing the reasonableness of warrantless intrusions that fall within recognized exceptions to the warrant requirement.

---

183. *See, e.g.*, Uviller, *supra* note 174, at 29-30. The generalized reasonableness approach is especially evident in the Court's drug testing decisions. *See infra* note 213.

184. *E.g.*, Harris v. United States, 331 U.S. 145, 196 (1946) (Jackson, J., dissenting) (the Framers "apparently . . . believed that by thus controlling search warrants they had controlled searches").

185. As Jacob Landynski, a leading search and seizure historian, pointed out a quarter of a century ago:

> The second clause . . . defines and interprets the first telling us the kind of search that is *not* "unreasonable," and therefore not forbidden, namely the one carried out under the safeguards there specified.
>
> . . . To detach the first clause from the second is to run the risk of making the second virtually meaningless.

LANDYNSKI, *supra* note 176 at 43-44.

Compare this to Professor Uviller's paraphrase of the position he attributes to Justice Stewart: "the [statement of the] right to be secure against *unreasonable* intrusions, followed immediately by a description of the means of obtaining a warrant, implies that the method described (search by warrant) is ordinarily the *reasonable* one." Uviller, *supra* note 174, at 33.

The traditional interpretation of Fourth Amendment reasonableness thus implicates two basic criteria for a reasonable intrusion. The first criterion is a substantive standard: a police intrusion can only be "reasonable" if it is supported by probable cause.[186] The second criterion is an allocation of decision-making authority for assessing the adequacy of the probable cause showing:[187] specifically, cause for an intrusion should normally be assessed by "a neutral and detached magistrate."[188] The traditional interpretation of Fourth Amendment reasonableness accords with the idea that the amendment is a statement of a citizen's right not to be subjected to government intrusions unless the government complies with specific substantive and procedural criteria—the probable cause standard and the warrant requirement.

Although it hardly can be said that the history of the Court's interpretation of the Fourth Amendment is free of confusion, the view that warrant intrusions should be taken as the model for reasonable searches clearly emerged as the dominant theoretical perspective by which the Court has addressed the Fourth Amendment.[189]

---

186. Probable cause necessarily includes a particularized expectation about what or who will be found where. The particularity language in the second clause of the Fourth Amendment should be understood to serve two distinct purposes. First, it specifies there cannot be probable cause unless there is particularity regarding the expectation of what will be found and where it will be found. Thus, whenever I write about probable cause I include this aspect of particularity in that reference. Second, the particularity requirement also serves to define the scope of the authority actually granted by a warrant. This second meaning applies only in the context of an issued warrant.

187. Though the point is so self-evident that it is often left implicit in discussions of the amendment, the Fourth Amendment is ultimately rooted in the concept of legality (as is the entire Constitution); government agents may not act unless they have been given legal authority to do so. The explicit statement of the requirements for a judicially issued warrant in the second (warrant) clause of the amendment should also be understood to express an allocation of legal decision-making authority as to who has the legal authority to authorize an intrusion into a citizen's privacy: the Framers expected that government agents generally could not possess legal authority to intrude on a citizen's privacy unless they were accorded that authority through a judicially issued warrant authorizing the specific intrusion. The Fourth Amendment's reasonableness requirement should be understood to demand that police intrusions be based on valid legal authority.

188. As Justice Jackson said:

> The point of the Fourth Amendment, which often is not grasped by zealous officers, is not that it denies law enforcement the support of the usual inferences which reasonable men draw from evidence. Its protection consists in requiring that those inferences be drawn by a neutral and detached magistrate instead of being judged by the officer engaged in the often competitive enterprise of ferreting out crime.

United States v. Johnson, 333 U.S. 10, 13-14 (1948). *See also* Wong Sun v. United States, 371 U.S. 471, 481-82 (1938).

189. *See* United States v. Lefkowitz, 285 U.S. 452 (1932); Go-Bart Importing

Thus, the Court adopted the view that a warrant was presumptively required for a reasonable intrusion but recognized there were exceptions to that presumption.[190] Justice Stewart[191] summed up this development in *Katz v. United States*:[192] "searches conducted outside the judicial process, without prior approval by judge or magistrate, are *per se* unreasonable under the Fourth Amendment—subject only to a few specifically established and well-delineated exceptions."[193] The Court continues to reiterate this statement as a "cardinal principle" of Fourth Amendment analysis.[194] Therefore, at least in theory, the

---

Co. v. United States, 282 U.S. 344 (1931); Agnello v. United States, 269 U.S. 20, 33 (1925) (even well-founded belief that evidence is concealed in a dwelling house does not justify a search without a warrant despite facts that unquestionably show probable cause); *see also* Jacob W. Landynski, *Comments on Uviller*, 25 CRIM. L. BULL. 51, 52 ("[T]he idea . . . that the Fourth Amendment requires a warrant except in unusual circumstances . . . has suffused search and seizure decisions from the very first.").

190. Justice Frankfurter expressed the traditional understanding as follows:

> The plain import [of the language used in the Fourth Amendment] is that searches are "unreasonable" unless authorized by a warrant, and a warrant hedged about by adequate safeguards. "Unreasonable" is not to be determined with reference to a particular search and seizure considered in isolation. . . . [W]ith minor and severely confined exceptions, inferentially a part of the Amendment, every search or seizure is unreasonable when made without a magistrate's authority expressed through a validly issued warrant.

United States v. Harris, 331 U.S. 145, 161 (1947) (Frankfurter, J., dissenting).

191. Justice Stewart is widely regarded as a proponent of this warrant-requirement-with-delineated-exceptions approach.

192. 389 U.S. 347 (1967).

193. *Id.* at 357.

194. *E.g.*, California v. Acevedo, 111 S. Ct. 1982 (1991); Mincey v. Arizona, 437 U.S. 385, 390 (1978). It must be noted, however, that the Court's reaffirmation of the cardinal principle announced in *Katz* may be more rhetorical than substantive. Professor Kamisar has noted the majority opinion in *Acevedo* contains the following:

> To the extent that the *Chadwick-Sanders* rule protects privacy, its protection is minimal. Law enforcement officers may seize a container and hold it until they obtain a search warrant . . . . 'Since the police, by hypothesis, have probable cause to seize the property, we can assume *that a warrant will be routinely forthcoming* in the overwhelming majority of cases.'

Yale Kamisar, *Arrest, Search and Seizure: Prepared Remarks at U.S. Law Week's Thirteenth Annual Constitutional Law Conference* 59 (1991) (mimeo) (quoting California v. Acevedo, 111 S. Ct. at 1987) (emphasis added by Kamisar). Professor Kamisar points out that:

> [I]t is hard to think of anything more anomalous than an opinion that contains both the language I just read and *another* passage reaffirming the "cardinal principle that 'searches conducted outside the judicial process, without prior approval by judge or magistrate, are *per se* unreasonable . . . .'"
>
> If the Court believes that a search warrant is a mere formality—and what else did it mean when it told us that whenever the police *believe* they possess probable cause "a warrant will be routinely forthcoming"? —*why*

Court restricts the settings in which warrantless searches can be made.[195]

In keeping with the warrant intrusion model, the Court has also understood probable cause to be a requirement for those recognized exceptions to the warrant requirement that apply to highly invasive searches. For example, the "automobile exception" that allows police to dispense with the warrant requirement for searches of automobiles is nevertheless predicated on the police being able to show probable cause to believe the automobile contains contraband or evidence.[196] Probable cause is also required for the "exigent circumstances" exception to the warrant requirement.[197]

2. The "Generalized Reasonableness" Reading of the Fourth Amendment

The generalized reasonableness reading of the amendment takes a very different view of the reasonableness requirement stated in the first clause of the amendment. This approach does not assume the Framers intended to adopt warrant searches as a model for a reasonable search.[198] Instead, proponents of this approach read the second clause in the amendment as though it were literally a warrant clause, as though the specific requirements set out in that clause were intended

---

> is it a cardinal principle that searches conducted outside the judicial process are "*per se* unreasonable"?

Kamisar, *supra*, at 60.

195. There is a clear discrepancy between theory and practice. The Court has created such a variety of exceptions to the warrant requirement and those exceptions apply to so many of the most common settings in which police intrusions occur that the number of constitutional warrantless intrusions made by the police greatly exceed the number of intrusions for which the police obtain a warrant. *See supra* note 17. The warrant requirement has long been something of a myth in practical terms. *See* Bradley, *supra* note 178, at 1473-74; James B. Haddad, *Well-Delineated Exceptions, Claims of Sham, and Fourfold Probable Cause*, 68 J. CRIM. L. & CRIMINOLOGY 198 (1977); Kamisar, *supra* note 176, at 567-71; Yackle, *supra* note 2, at 414-15.

196. The automobile exception to the warrant requirement was announced in Carroll v. United States, 267 U.S. 132 (1925), which made probable cause a condition for that exception. *Id.* at 153, 161 (noting "the necessity of probable cause"). *See also* Chambers v. Maroney, 399 U.S. 42, 51 (1970) (probable cause is a minimum requirement for a constitutional search of an automobile). The Court continues to adhere to the probable cause requirement for automobile searches under the automobile exception (at least outside the inventory search setting). *See* California v. Acevedo, 111 S. Ct. 1982 (1991).

197. Warden v. Hayden, 387 U.S. 294, 298-99 (1967).

198. For commentary supporting this reading see Akhil Reed Amar, *The Bill of Rights as a Constitution*, 100 YALE L.J. 1131, 1178-80 (1991); Richard A. Posner, *Rethiniking the Fourth Amendment*, 1981 SUP. CT. REV. 49, 72-73; TAYLOR, *supra* note 180. *But see* Grano, *supra* note 176, at 619-20.

only to prevent the issuance of warrants comparable to general warrants of the sort used by British officials in the colonial period.[199] The limited meaning accorded to the second clause makes it possible to read the "reasonableness" requirement in the first clause as a free-standing, generalized, and flexible standard for constitutional intrusions.

As a result, the generalized reasonableness reading rejects both the idea that there should be a warrant requirement[200] and the idea that the probable cause standard for the issuance of warrants should be viewed as a general benchmark for assessing the reasonableness of warrantless police intrusions.[201] Additionally, the generalized reasonableness reading allows for a flexible, colloquial interpretation of "reasonableness" under which any police conduct that is "understandable" in the circumstances according to common sense should be judged "reasonable" for purposes of assessing the constitutionality

---

199. There is broad agreement that such general warrants were viewed as an evil by the Framers. *See, e.g.*, United States v. Chadwick, 433 U.S. 1, 7-8 (1977) ("It cannot be doubted that the Fourth Amendment's commands grew in large measure out of the colonists' experience with the writs of assistance and their memories of the general warrants formerly in use in England.").

200. Justice Scalia recently has endorsed the generalized reasonableness reading. *See infra* note 217. He also has recently suggested the Framers' purpose in specifying the requirements for warrants was only "to preserve the jury's role in regulating searches and seizures" through civil trespass actions. California v. Acevedo, 111 S. Ct. 1982, 1992 (1991) (Scalia, J., concurring) (citing Amar, *The Bill of Rights as a Constitution*, 100 YALE L.J. 1131, 1178-80 (1991)). He explains a warrant provided absolute personal immunity to the officer who conducted an intrusion, but that there was no immunity for warrantless intrusions. 111 S. Ct. at 1992. *But see* Dellinger, *Of Rights and Remedies: The Constitution as a Sword*, 85 HARV. L. REV. 1532, 1540 (1972).

The introduction of a fresh understanding of the Framers' intent after nearly eighty years of Supreme Court interpretation of the Fourth Amendment poses some difficulties. For example, the claim that the Framers were concerned with the preservation of trespass actions raises questions as to the constitutionality of the Court's recent expansion of immunity for officers who conduct warrantless searches. *E.g.*, Anderson v. Creighton, 493 U.S. 635 (1987) (extending immunity to officers who conduct unconstitutional, warrantless searches based on reasonable, good-faith mistake about search law). As Justice Scalia concedes, "elimination of the common law rule that reasonable, good-faith belief was no defense to absolute liability for trespass . . . may make a warrant indispensible to reasonableness where it once was not." *Acevedo*, 111 S. Ct. at 1993.

201. The generalized reasonableness interpretation can be criticized on textual grounds. It seems improbable that the Framers would have gone to the trouble of specifying standards for warrants if they did not think the government at least usually would be required to obtain a warrant before intruding on a citizen's privacy. Facially, the Fourth Amendment is considerably more detailed than other provisions of the Bill of Rights. The inclusion of relatively detailed statements regarding the requirements for valid warrants in a constitutional provision would appear to indicate the Framers expected those requirements to be of broad significance. *See* Kamisar, *supra* note 176, at 575.

of police intrusions.[202] In effect, this reading treats the Fourth Amendment more as a "regulatory canon" than as a statement of a citizen's enforceable right.[203]

3. The Triumph and Decline of the Traditional Understanding

Although the Court flirted with the generalized reasonableness reading of the Fourth Amendment in some of its early search cases,[204] it appeared the generalized reasonableness interpretation clearly had lost out to the traditional understanding of the amendment articulated in Justice Stewart's opinion in *Katz*. The generalized reasonableness reading did not, however, disappear. In fact, it was given new life by the Court's recognition of an unprecedented exception to the warrant requirement in *Terry v. Ohio*[205] only one year after *Katz*.

---

202. Professor Uviller describes the generalized reasonableness reading as a "pure reasonability" approach. Uviller, *supra* note 174, at 34-35. Compare the generalized reasonableness interpretation to Professor Bradley's description of the "no lines" model of the Fourth Amendment. Bradley, *supra* note 178, at 1489-91.

203. The most significant drawback of the generalized reasonableness reading is simply that its notion of reasonableness is virtually devoid of content. Its colloquialized notion of reasonableness operates more as a conclusory label than a criterion. As a result, it is difficult to square the generalized reasonableness reading even with the widely accepted view that the Framers were especially concerned with the abuses associated with the use of general warrants during the colonial period. Viewed functionally, the evil in a general warrant was that it authorized a colonial official to intrude into subjects' homes at the offical's own discretion even if the official did not have particularized cause. TAYLOR, *supra* note 180, at 30 (the evil that was perceived in the general warrant was the absence of a requirement of a specific cause for the search). If today's police officer is allowed to intrude on citizens' privacy without obtaining a warrant based on probable cause for that particular search—and particularly if he is allowed to make a warrantless intrusion without being subjected at least to the probable cause standard—then the officer would have the same sort of unstructured discretion as he would under a general warrant. It seems unlikely the Framers would have accepted that the government can bestow generalized discretionary authority on a police officer through the credential of a metal police badge (backed up, of course, by statutory authority) when they clearly would not have allowed the same officer to be given the same generalized discretionary authority in the form of a paper general warrant.

204. *See, e.g.*, United States v. Rabinowitz, 339 U.S. 56 (1950). In *Rabinowitz*, Justice Minton wrote:

> What is a reasonable search is not to be determined by any fixed formula. The Constitution does not define what are "unreasonable searches," and regrettably in our discipline we have no ready litmus-paper test . . . . The relevant test is not whether it is reasonable to procure a search warrant, but whether the search was reasonable. That criterion in turn depends on the facts and circumstances—the total atmosphere of the case.

339 U.S. at 63, 66 (1950); *see also* Marron v. United States, 275 U.S. 192 (1927); Jacob W. Landynski, *Comments on Uviller*, 25 CRIM. L. BULL. 51, 52-53 (1989).

205. 392 U.S. 1 (1968). I oversimplify this history in an effort to be succinct without being misleading. For example, the balancing approach in *Terry* is also

In *Terry* the Court recognized that an intrusion that is less invasive of privacy than an arrest could be made, not only without a warrant, but also on a lesser showing of cause than probable cause. Although *Terry* itself retained a requirement of individualized cause,[206] *Terry* was unsettling because it introduced a balancing approach to "reasonableness" under which the standard for a reasonable intrusion could be adjusted to reflect the degree to which the intrusion invaded the citizen's privacy.[207] This endorsement of a balancing approach gave rise to claims of "reasonableness" that fell outside of the traditional understanding of Fourth Amendment reasonableness.[208] In

---

evident in the Court's treatment of administrative searches a year earlier. *See* Camara v. Municipal Court, 387 U.S. 523, 538 (1967); See v. City of Seattle, 387 U.S. 541, 545 (1967).

For commentary on this development, see Ronald Bacigal, *The Fourth Amendment in Flux: The Rise and Fall of Probable Cause*, 1979 U. ILL. L. F. 763; Wayne R. LaFave, *Administrative Searches and the Fourth Amendment: The* Camara *and* See *Cases*, 1967 SUP. CT. REV. 1; Scott E. Sundby, *A Return to Fourth Amendment Basics: Undoing the Mischief of* Camara *and* Terry, 72 MINN. L. REV. 383 (1988); White, *supra* note 53, at 176-80.

206. *Terry* concluded a warrantless pat-down search would be constitutional as long as the officer had at least articulable suspicion—a lower standard of cause than probable cause (but still a standard of cause). Chief Justice Warren emphasized, however, that "[t]his demand for specificity in the information upon which police action is predicated is the central teaching of this Court's Fourth Amendment jurisprudence." 392 U.S. at 21 n.18. The *Terry* Court concluded pat-down searches would be reasonable only if "the police officer [is] able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant [the particular] intrusion." *Id.* at 21. Immediately before stating this standard of articulable cause, Chief Justice Warren wrote "the notions which underlie both the warrant procedure and the requirement of probable cause remain fully relevant in this context." *Id.* at 20. It is evident *Terry* was written in the hope it would not lead to a full-blown generalized reasonableness approach to assessing the reasonableness of police intrusions.

207. *Terry* speaks of "balancing the need to search [or seize] against the invasion which the search [or seizure] entails." 392 U.S. at 21 (brackets in original). Reference to the need for a search is a reference to the cause for a search. There clearly cannot be a need to search in the absence of cause to do so. In contrast, more recent opinions employing a generalized reasonableness approach tend to advocate balancing in terms of governmental interests without any reference to particularized need. *E.g.*, New Jersey v. T.L.O., 469 U.S. 325, 341 (1985); United States v. Place, 462 U.S. 696, 704 (1983); Delaware v. Prouse, 440 U.S. 648, 654 (1979).

208. Ironically, it appears the development of a lower standard of cause for less invasive searches in *Terry* was stimulated (perhaps even necessitated) by the Warren Court's decision to give an expansive meaning to the scope of Fourth Amendment protections in *Katz v. United States*, 389 U.S. 347 (1967). In effect, the Court decided it would not attempt to set a threshold as to how invasive a governmental intrusion had to be before it would become a search or seizure subject to the Fourth Amendment. Thus, the Court decided in *Terry* that even a pat-down or frisk of a suspect was a search. 392 U.S. at 16. This decision brought a much greater sphere of police-citizen contacts under the Fourth Amendment. The indirect

effect, *Terry* opened the door for the application of a generalized reasonableness approach to less-invasive intrusions.[209] Subsequently, the Burger Court and Rehnquist Court have seized on that opening to declare a wide variety of purportedly less invasive government intrusions to be "reasonable."[210] Recent decisions have applied a balancing approach to a wide array of government intrusions that involve "minimal" invasions of privacy in a criminal law enforcement context[211] or "special" government "needs" in nonpolice intrusions (some of which are quite invasive).[212] The Court even has held to be "reasonable" some intrusions that do not rest on any form of individualized suspicion.[213] These developments have given renewed impetus to the generalized reasonableness interpretation,[214] and it is

---

effect of taking an expansive approach to the scope of Fourth Amendment protection was the unleashing of a pragmatic demand that less invasive intrusions be allowed without requiring warrants or probable cause.

209. *See* Silas J. Wasserstom, *The Court's Turn Toward A General Reasonableness Interpretation of the Fourth Amendment*, 27 AM. CRIM. L. REV. 119 (1989); Silas J. Wasserstrom, *The Incredible Shrinking Fourth Amendment*, 21 AM. CRIM. L. REV. 257 (1984).

210. The only predictable aspect of balancing tests is that the outcome will be determined by the identity of the balancers.

211. *E.g.*, Michigan v. Long, 463 U.S. 1032, 1045-52 (1983) (approving a *Terry* type of frisk that stretched into a warrantless search of an automobile interior despite the absence of probable cause). *See also* Wasserstrom, *supra* note 209.

212. The Court has reduced the requirements for reasonable searches in the administrative and regulatory context. *E.g.*, Marshall v. Barlow's, Inc., 436 U.S. 307, 320-21 (1978). The Court also has adopted a "special needs" approach to nonpolice searches. *E.g.*, O'Connor v. Ortega, 480 U.S. 709 (1987) (constitutionality of search of government worker's office by supervisor should be assessed in terms of reasonableness under all the circumstances); Griffin v. Wisconsin, 483 U.S. 868 (1987) (special needs of supervising probationers makes it reasonable to depart from the usual warrant and probable cause requirements and to allow warrantless search of probationer's home on reasonable suspicion); New Jersey v. T.L.O., 469 U.S. 325 (1985) (constitutionality of search of student's purse by school administrator should be assessed in terms of reasonableness under all the circumstances).

213. The Court initially abandoned individualized suspicion when it ruled border guards could stop vehicles at highway checkpoints near the border to detect illegal aliens without individualized cause to suspect the presence of aliens. United States v. Martinez-Fuerte, 428 U.S. 543 (1976). Subsequently, the Court has ruled that cars may be stopped to identify inebriated drivers at highway checkpoints, under certain conditions, regardless of the absence of individualized suspicion. Michigan Dept. of State Police v. Sitz, 110 S. Ct. 2481 (1990). *See also* Skinner v. Railway Labor Executives' Ass'n, 109 S. Ct. 1402 (1989) (upholding drug testing of railway crews involved in train accidents in the absence of individualized suspicion of drug use); National Treasury Employers Union v. Von Raab, 109 S. Ct. 1384 (1989) (upholding drug testing of Customs Service employees holding certain positions in the service in the absence of individualized suspicion of drug use).

214. Professor Uviller has attempted to quantify the shift in the Court's approach by classifying the style of opinions in warrantless search cases as "categorical" (my traditional interpretation) or "reasonability" (my generalized reasonableness approach) in the years immediately before and after Justice Stewart's retirement.

apparent some Justices (including Chief Justice Rehnquist[215] and Justices White[216] and Scalia[217]) often find it the more attractive way to read the amendment.

The emergence of generalized reasonableness since *Terry* has had a major influence on Fourth Amendment law. It appears the balancing approach endorsed by generalized reasonableness is now the dominant approach in the Court's decisions in *warrantless* searches.[218] One result of the prominence of generalized reasonableness is that there is a good deal of confusion regarding the criteria for such searches.[219] In fact, much of the recent commentary condemning the confused state of Fourth Amendment doctrine addresses developments that can be traced to *Terry*.[220]

He reports the ratio of categorical to reasonability opinions from 1975 to 1981 was twenty-three to eight, but that it was ten to nineteen in 1984 to 1988. Uviller, *supra* note 174, at 38.

215. Then-Justice Rehnquist appeared to endorse generalized reasonableness in his majority opinion in Cady v. Dombrowski, 413 U.S. 433 (1973):

> The Framers of the Fourth Amendment have given us only the general standard of "unreasonableness" as a guide in determining whether searches and seizures meet the standard of that Amendment in those cases where a warrant is not required . . . . [A]nd very little that we might say here can usefully refine the language of the Amendment itself in order to evolve some detailed formula for judging cases such as this.

Id. at 448. (citations omitted). Note this statement does not apply to intrusions requiring a warrant. Justice Rehnquist also adopted a novel interpretation of reasonableness in *Robinson*. *See supra* note 169.

216. Justice White has been widely viewed as the leading exponent of the generalized reasonableness reading on the Court. He advocated "the commonsense standard of reasonableness governing search and seizure cases" in Coolidge v. New Hampshire, 403 U.S. 443, 527 (1971) (concurring opinion). *See also* Uviller, *supra* note 174, at 34-36; Landynski, *supra* note 189, at 52; *see generally*, Roger B. Dworkin, *Fact Style Adjudication and the Fourth Amendment: The Limits of Lawyering*, 48 IND. L.J. 329 (1973).

217. Justice Scalia also appears to be emerging as an advocate of the generalized reasonableness reading. Subsequent to *Rodriguez*, he wrote "the Fourth Amendment does not by its terms require a prior warrant for searches and seizures; it merely prohibits searches and seizures that are 'unreasonable.' What it explicitly states regarding warrants is by way of limitation upon their issurance rather than requirement of their use." California v. Acevedo, 111 S. Ct. 1982, 1992 (1991) (Scalia, J., concurring) (citation omitted). This is a pure statement of generalized reasonableness. Justice Scalia, however, qualified this statement by acknowledging that warrants should be required "where the common law required a warrant." *Id.* at 1993; *see also supra* note 200.

218. *See generally*, Uviller, *supra* note 174; Wasserstrom, *supra* note 209.

219. Professor Amsterdam has suggested the "sliding-scale" produced by the generalized reasonableness approach turns Fourth Amendment law into "one immense Rorschach blot." Amsterdam, *supra* note 7, at 393.

220. For a collection of complaints regarding the confused state of Fourth Amendment law, see Bradley, *supra* note 178, at 1468-69. Bradley describes the state of Fourth Amendment law as a "tarbaby: a mass of contradictions and obscurities that has ensnared the 'Brethren' in such a way that every effort to extract themselves

The general state of doctrinal confusion, however, should not be allowed to obscure the more important practical implication of generalized reasonableness: the government usually wins under this approach.[221] The absence of anything approaching an operational standard for determining the constitutionality of an intrusion under the generalized reasonableness reading amounts to an invitation to reviewing courts to treat a police intrusion as "reasonable" if any explanation for the police conduct can be given. A regime of unstructured "reasonableness" hardly makes citizens "secure in their persons, houses, papers, and effects," or gives them a right of much consequence. The absence of clear standards for "reasonableness" trivializes Fourth Amendment rights.[222]

The ascendancy of the generalized reasonableness approach since *Terry*, however, has not been pervasive. The balancing approach called for the retention of both the probable cause requirement and the presumptive warrant requirement in highly invasive intrusions because of the weight that must be accorded the citizen's privacy interests in such intrusions. Therefore, even Burger and Rehnquist Court decisions have refrained from applying a generalized reasonableness analysis to the most highly invasive police intrusions, such as police intrusions of homes. In that area, the traditional understanding of Fourth Amendment reasonableness appeared to be vital.

---

only finds them more profoundly stuck." *Id.* at 1468. He also complains "the Court has iffed, anded, and butted the fourth amendment into hopeless confusion. . . ." *Id.* at 1482; *see also* Silas J. Wasserstrom & Louis Michael Seidman, *The Fourth Amendment as Constitutional Theory*, 77 GEO. L.J. 19 (1988).

Not all commentators are critical of the Court's approach, however. For less critical accounts of these developments, see Wayne R. LaFave, *Being Frank About the Fourth: On Allen's "Process of 'Factualization' in the Search and Seizure Cases,"* 85 MICH. L. REV. 427 (1986); Uviller, *supra* note 174, at 37.

221. The increasing influence of the generalized reasonableness approach appears to be linked to an increasing tendency for the Court to come to progovernment decisions. Professor Uviller reports the Court only upheld 25.8% of the warrantless searches it reviewed from 1975 to 1981 (when Justice Stewart acted as a spokesman for the traditional interpretation) but the Court upheld 76.3% of the warrantless searches it reviewed from 1982 to 1988 (after Justice Stewart's retirement). Uviller, *supra* note 174, at 38, App. 1.

As Uviller (who is not disturbed by the rise of generalized reasonableness) explains, "the circumstantial test of reasonability may net more affirmances [of warrantless police searches] than the crabbed hunt through the lawbooks for a previously labeled [exception to the warrant requirement] into which the police conduct in question neatly fits." *Id.* at 37.

222. *See supra* note 9; Yackle, *supra* note 2, at 336. Not all legal commentators share the view that generalized reasonableness necessarily implies lessened privacy protection. For example, Professor Bradley has argued the choice between a "bright lines" approach to the Fourth Amendment (the traditional interpretation) and a "no lines" approach (generalized reasonableness) is neutral in terms of results. Bradley, *supra* note 178, at 1501.

As a result, majorities in the Burger and Rehnquist Courts have continued to espouse the two basic principles of the traditional understanding of Fourth Amendment reasonableness when police intrusions of homes were under review. First, the Court has continued to insist intrusions of homes must be authorized by warrant except where genuine exigencies are present.[223] Second, the Court has continued to require probable cause even in those instances in which a warrantless intrusion of a residence is allowed under a recognized exception to the warrant requirement.[224] The latter principle was reasserted, for example, in Justice Scalia's majority opinion in the 1987 *Arizona v. Hicks* decision.[225]

There was strong dissent, however, in *Hicks* and in the other recent cases that endorsed the traditional understanding of Fourth Amendment reasonableness in intrusions of residences.[226] It now appears a majority of the Justices are willing to apply a generalized reasonableness analysis to uphold even highly invasive warrantless police intrusions of homes. At least this is what *Rodriguez* portends. Justice Scalia's majority opinion in *Rodriguez* invokes a colloquialized notion of "reasonableness" to uphold an extremely invasive intrusion of a home. In doing so it dispenses not only with the warrant requirement but also with the requirement that police officers must

---

223. *See* Skinner v. Railway Labor Executive Ass'n., 489 U.S. 602, 622 (1989) (a warrant is the baseline for reasonableness of a search or seizure in the home); Arizona v. Hicks, 480 U.S. 321, 326-27 (1987) (warrantless searches and seizures in a home are presumptively unreasonable); Welsh v. Wisconsin, 466 U.S. 740 (1984) (arrest in home unconstitutional in absence of a warrant); United States v. Karo, 468 U.S. 705, 714-15 (1984); Payton v. New York, 445 U.S. 573, 586 (1980); Mincey v. Arizona, 437 U.S. 385, 394 (1978) (warrantless searches of residences are permissible only if compelling law enforecement purpose involved); Michigan v. Tyler, 436 U.S. 499, 509 (1978) (warrantless entry of burning building justified only because of compelling need for official action and no time to secure a warrant).

The Court's insistence that warrants be obtained for intrusions of homes is based on the recognition that "physical entry of the home is the chief evil against which the wording of the Fourth Amendment is directed." United States v. United States District Court, 407 U.S. 297, 313 (1972); *see also* Steagald v. United States, 451 U.S. 204, 211 (1981); Silverman v. United States, 365 U.S. 505, 511 (1961).

224. Probable cause is required for a warrantless police entry of a home on the basis of exigent circumstances. *See* Warden v. Hayden, 387 U.S. 294, 307 (1967).

225. 480 U.S. 321. Justice Scalia wrote for the majority "[a] dwelling-place search, no less than a dwelling-place seizure, requires probable cause . . . ." *Id.* at 328. Lower courts have read *Hicks* as a reaffirmation of the probable cause requirement for any intrusion of a home. *E.g.*, United States v. Howard, 828 F.2d 552, 555 (9th Cir. 1987) (citing *Hicks* as authority that "entry into a person's home is so intrusive that such searches always require probable cause regardless of whether some exception would excuse the warrant requirement").

226. *Hicks*, 480 U.S. at 335-39 (O'Connor, J., dissenting) (arguing probable cause should not be required for minimally intrusive police conduct occuring within a home).

have probable cause (in fact it dispenses with cause entirely) for the intrusion.

B. Rodriguez*'s Downgraded Notion of "Reasonableness"*

As discussed above, Justice Scalia obtains maneuvering room to construct his rationale in *Rodriguez* by revising the Court's understanding of the nature of consent and its relationship to Fourth Amendment reasonableness.[227] He begins by adopting the novel posture that consent is an exception to the warrant requirement.[228] He then transforms that statement into a claim that "the consent of the person or his cotenant [is one of the] various elements . . . that can make a search of a person's house 'reasonable.'"[229] He further transforms that characterization of consent into a claim that consent is a "factual determination" and asserts the "reasonableness" of any such determination should be assessed *ex ante* from the vantage point of the police officer.[230] On that basis, he advances that "seeming consent," an appearance that a "consenting" person has authority to consent, is enough because "[i]t is apparent that in order to satisfy the 'reasonableness' requirement of the Fourth Amendment, what is generally demanded of the many factual determinations that must regularly be made by agents of the government . . . is not that they always be correct, but that they always be reasonable."[231] He concludes, therefore, a police intrusion is "reasonable" and satisfies the Fourth Amendment even if it is based only on "seeming consent."[232]

Justice Scalia's rationale strays a long way from the traditional understanding of Fourth Amendment reasonableness. The most fundamental defect in his formulation is the total omission of the core requirement of Fourth Amendment reasonableness: a showing of probable cause for the intrusion.

1. *Rodriguez*'s Omission of Any Assessment of Probable Cause

The traditional understanding of Fourth Amendment reasonableness insists any entry of a residence that purportedly falls within an exception to the warrant requirement must still rest on a showing of

---

227. *See supra* text accompanying notes 135-36.
228. *See supra* text accompanying notes 137-40.
229. *Rodriguez*, 110 S. Ct. at 2799.
230. *See supra* note 136.
231. *Rodriguez*, 110 S. Ct. at 2800.
232. In fact, if consent searches are to be assessed in terms of "reasonableness," then "seeming consent" is all that is relevant. Whether consent is actually valid is irrelevant to a "reasonableness" claim.

probable cause.[233] This cause requirement is the core idea in the Fourth Amendment and the essence of the right it proclaims.[234] The cause requirement denies government agents any authority to intrude into a citizen's privacy merely because they want to. Rather, the cause requirement permits an intrusion only if the government agents can provide a specific kind of justification for undertaking an intrusion of a citizen's residence: government agents must have a substantial expectation, grounded in credible information, about what the intrusion will produce. For a search, the government must have information that creates a substantial expectation that a *particular* item or items of evidence relating to a *particular* crime will be found in the *particular* residence to be searched.[235] For an arrest, the government must have information creating a substantial expectation that a *particular* offense has been committed by the *particular* person to be arrested.[236]

The most striking feature of Justice Scalia's rationale is his omission of any assessment of probable cause in a "seemingly consented" intrusion. Neither genuine consent nor "seeming consent" provides probable cause (or any degree of cause whatsoever) for an intrusion. Consent to an intrusion does not create any expectation that an intrusion will produce anything in particular, or even anything at all. Even genuine consent provides only permission for an intrusion; it does not provide cause for one, so it cannot satisfy Fourth Amendment reasonableness. "Seeming consent" does not provide either permission or cause; therefore, it clearly cannot satisfy the traditional understanding of Fourth Amendment reasonableness.

---

233. *See supra* text accompanying notes 223-25.

234. *Cf.* Chambers v. Maroney, 399 U.S. 42, 51 (1970) (probable cause is the "minimum requirement for a reasonable search permitted by the Constitution"); Terry v. Ohio, 392 U.S. 1, 21 n.18 (1968) ("The demand for specificity in the information upon which police action is predicated is the central teaching of the Court's Fourth Amendment jurisprudence.").

235. My brief discussion necessarily glosses over a significant expansion of police power beyond that envisioned by the Framers: the jetisoning of the prohibition against searches for and seizures of mere evidence. *See* Zurcher v. Stanford Daily, 436 U.S. 547, 577-79 (1978) (Stevens, J., dissenting); Warden v. Hayden, 387 U.S. 294 (1967).

236. The showing of cause required for arrest of a suspect in a residence varies depending on whether the residence involved is the suspect's own or that of a third party. Payton v. New York, 445 U.S. 573 (1980), allows police to enter the suspect's own home for the purpose of arresting the suspect based on an arrest warrant that is issued on the basis of probable cause to believe the suspect has committed a particular crime. The Court declined to require a search warrant, based on probable cause to believe the subject will be found at the dwelling, in this setting. *Id.* at 602-03. An arrest warrant also suffices for an entry to arrest a resident guest. *See* Minnesota v. Olson, 495 U.S. 91 (1990).

For an entry of a third-party's residence to effectuate arrest of a nonresident suspect, however, the police must obtain a search warrant based on probable cause to believe the suspect will be found in the third-party's residence. *See* Steagald v. United States, 451 U.S. 204 (1981).

The rationale for *Rodriguez*, in other words, does not amount simply to the announcement of a new exception to the warrant requirement within the parameters of the traditional understanding of Fourth Amendment reasonableness.[237] Instead, Justice Scalia's claim that "seeming consent" can make an intrusion "reasonable" ultimately treats the reasonableness standard in the first clause of the amendment as though it were merely an invocation of a colloquial notion of reasonableness that does not necessarily include an assessment of cause. *Rodriguez*, in other words, rests on an extreme version of generalized reasonableness.

Justice Scalia, however, does not acknowledge his adoption of the generalized reasonableness interpretation of the Fourth Amendment. He is not candid about the unprecedented nature of his application of that interpretation to an extremely invasive intrusion into a residence. To the contrary, he employs a variety of rhetorical diversions and obfuscations that camouflage the majority's repudiation of the traditional understanding of Fourth Amendment reasonableness.

The first rhetorical device that Justice Scalia employs is his reference to consent as one of "various elements" that make searches "reasonable." Use of the term "various elements" conveys an impression that "reasonableness" is aconceptual and directs attention away from the traditional notion that Fourth Amendment reasonableness is an analytic concept with a specifiable meaning. "Various elements" implies "reasonableness" is merely a grab-bag of idiosyncratic judicial choices about what kinds of police conduct are good for society and what kinds are not. The rhetorical advantage offered by the various elements nomenclature is apparent: it is much easier to justify the addition of a new element to a grab-bag than it is to justify an incongruent application of a conceptual structure.

Justice Scalia follows up this deconceptualizing move by passing off consent (or, actually, "seeming consent") as though it were just another element in the grab-bag of "reasonableness." He treats consent as though it were one of the recognized exceptions to the warrant requirement even though he does not cite any prior decision in which the Court actually treated it that way (none ever has).[238] He also fails to disclose that each of the recognized exceptions he mentions do require probable cause.[239]

Justice Scalia also imparts a gloss of continuity to his aconceptual treatment of Fourth Amendment reasonableness by mentioning the cause requirement of the traditional understanding of Fourth Amendment reasonableness while being careful not to accord it too much significance. In a curious but pregnant passage he writes:

---

237. Justice Marshall's dissenting opinion appears to interpret *Rodriguez* in this fashion. *See* 110 S. Ct. at 2803-04.

238. *See supra* notes 137-50 and accompanying text.

239. *See infra* notes 290-91.

> The fundamental objective that alone validates *all unconsented* government searches is, of course, the seizure of persons who have committed or are about to commit crimes, or of evidence related to crimes. But "reasonableness" with regard to this *necessary element*, does not demand that the government be factually correct in its assessment that that is what a search will produce. Warrants need only be supported by probable cause . . . .[240]

This passage obliquely acknowledges the probable cause requirement. The probable cause requirement is what Justice Scalia must be referring to when he writes of the "necessary element" of an "objective" that entails an "assessment [of what] a search will produce."[241] He is careful, however, not to use the term "probable cause" except in the specific context of the issuance of a warrant.[242]

Justice Scalia, moreover, writes as though this necessary element of an objective applied only to unconsented searches. He offers no reason why that should be so. If consent is just one of the various elements that can justify a search of a home, why does it not exhibit the *necessary characteristic* required of all of the other (unconsented) elements? Why should Fourth Amendment reasonableness have alternative meanings depending on whether the intrusion at issue is consented or unconsented?

It is easy enough to trace the reason that the Court did not apply any cause requirement in prior cases that approved consent intrusions. Those decisions were premised on the understanding that valid consent makes the Fourth Amendent and the probable cause requirement inapplicable.[243] This reason, however, is not available under Justice Scalia's novel treatment of consent as an element that satisfies Fourth Amendment reasonableness. How can "seeming consent" make a search "reasonable" under the Fourth Amendment if "seeming consent" does not exhibit the necessary condition for reasonableness? Justice Scalia does not explain why the probable cause requirement should not apply to the element of consent; he simply asserts the distinction. *Rodriguez* ultimately rests on an unexplained, arbitrary ejection of the probable cause requirement—the substantive core of the Fourth Amendment—from the colloquial notion of "reasonableness."

---

240. 110 S. Ct. at 2799 (emphasis added).

241. This statement not only is unnecessarily indirect, but also is deficient in its apparent unwillingness to admit that the probable cause standard is generally applicable to warrantless searches as well as warrant searches.

242. This containment of any reference to probable cause to the warrant setting is consistent with the "generalized reasonableness" reading of the amendment discussed above. *See supra* text accompanying notes 198-203.

243. For a discussion of the traditional concept of consent as a waiver or foregoing of protected privacy, see *supra* notes 89-97 and accompanying text.

Because he lacks an explanation for this move, Justice Scalia excuses his position by downplaying the significance of the standard he rejects. He notes that probable cause, as defined in *Illinois v. Gates*,[244] "demands no more than a proper assessment of probabilities in particular factual contexts."[245] This is true, of course. It even may be true that the *Gates* majority's adoption of a "totality of the circumstances" approach to probable cause and its description of probable cause as a "fair probability"[246] effectively eviscerates that standard.[247] The fact that probable cause means less after *Gates* than it did before is, however, no reason to deem a consent intrusion of a residence "reasonable" when it does not even meet the relaxed showing of cause allowed by *Gates*.

2. *Rodriguez*'s Evasion of the Warrant Requirement

Justice Scalia's treatment of the warrant requirement is equally arbitrary. He writes: "Another element often, though not invariably, required in order to make an *unconsented* search 'reasonable' is, of course, that the officer be authorized by a valid warrant."[248] This treatment again asserts a distinction between the meaning of reasonableness in consented and unconsented searches. Again, it begs the question of why consented intrusions should not be held to the same reasonableness standard to which unconsented searches are held. The Court did not require warrants for consent searches in past decisions because it viewed the Fourth Amendment and its warrant requirement as inapplicable if valid consent was given.[249] It is hardly self-evident, however, that consent searches should not require magistrate approval if consent is viewed as an element that satisfies the Fourth Amendment. If third-party authority to consent amounts to a "factual determination," as Justice Scalia insists,[250] and if consent is to be treated as though it could be a substitute for cause (which seems to be how Justice Scalia treats it), then it is not apparent why it should be any less appropriate to submit a determination regarding third-party authority to consent to a magistrate than it is to submit a determination of probable cause.

---

244. 462 U.S. 213 (1983).

245. *Rodriguez*, 110 S. Ct. at 2799 (citing Illinois v. Gates, 462 U.S. 213, 232 (1983)).

246. 462 U.S. at 238, 246.

247. Justice White expressed concern that *Gates* threatened to eviscerate probable cause. *Gates*, 462 U.S. at 272 (White, J., concurring).

248. 110 S. Ct. at 2799 (emphasis added).

249. *See supra* notes 93-97 and accompanying text.

250. I do not mean to suggest consent is accurately characterized as merely a factual determination. *See infra* text accompanying notes 300-12. I adopt Justice Scalia's characterization only to explore its implications.

3. *Rodriguez*'s Adoption of a Colloquialized Notion of "Reasonableness"

Justice Scalia does not make any attempt to justify his treatment of "seeming consent" within the terms of a traditional analysis of Fourth Amendment reasonableness. He does not dispute that consent does not satisfy either the probable cause requirement or the warrant requirement (it is beyond argument that it does not). He only asserts those criterion do not matter in "consented searches." If that is so, what does "reasonableness" mean in *Rodriguez*?

Justice Scalia writes that "What the defendant is assured by the Fourth Amendment is . . . that no [search of his home] will occur unless it is reasonable."[251] This is true, but it begs the crucial question: what are the criteria for determining whether an intrusion is "reasonable"? Justice Scalia does not supply any; he simply repeats the words "reasonable" and "reasonableness" as though they were self-defining and irreducible.[252] The only guidance he provides is one passage that appears to equate "reasonable" and "responsible" police conduct,[253] and another passage that suggests Fourth Amendment reasonableness is satisfied whenever police conduct is "understandable" in the situation.[254] The essence of Justice Scalia's position seems to be that Fourth Amendment reasonableness does not hold any more precise meaning than the loose meaning the word "reasonable" has in everyday speech. The problem with his position is simply that "reasonable" is a very elastic term in ordinary speech. The detrimental implications of this trivialization of constitutional language and doctrine bear close attention.

C. Why the Absence of a Probable Cause Requirement for "Seemingly Consented Searches" Matters

The most basic failing in Justice Scalia's claim that "seemingly consented searches" are "reasonable" is his application of that label without addressing whether the police have cause (a substantial reason) for wanting to undertake an intrusion.[255] This cause inquiry should

---

251. 110 S. Ct. at 2799.

252. For example, he writes there is no reason "why seemingly consented searches are 'unreasonable,' which is all that the Constitution forbids." *Id.* at 2800 n.*.

253. *Id.*

254. *Id.* at 2800 (quoting Hill v. California, 401 U.S. 797, 803-04). The language Justice Scalia quotes occurs in a discussion of the probable cause standard. *See infra* notes 318-26 and accompanying text.

255. Justice Marshall's dissenting opinion criticizes the majority opinion primarily for failing to take account of the warrant requirement. This approach, however, puts the allocation of decision-making authority ahead of the more essential substantive standard, probable cause, for the decision to be made.

be crucial in assessing Fourth Amendment reasonableness. As Justice Scalia correctly observed in *Hicks*, "[d]ispensing with the need for a warrant is worlds apart from permitting a lesser standard of *cause* . . . than a warrant would require, i.e., the standard of probable cause."[256] In *Rodriguez*, however, Justice Scalia bypasses the cause inquiry entirely and asks merely whether the police had reason to think a third party possessed authority to consent. In other words, *Rodriguez* evades the basic question of whether the police had reason to seek consent in the first place.[257]

Justice Scalia's omission of any cause inquiry suggests police requests for consent to an intrusion fall under the rubric that "it never hurts to ask." The Court's subsequent decision in *Florida v. Bostick*[258] demonstrates the Court does indeed take that view of requests for consent. In *Bostick* the Court approved an intrusion in which police officers found drugs after they boarded a bus, looked over the seated passengers, singled out Bostick (either randomly or because they "suspect[ed] in some vague way" that he might be involved in drugs[259]), approached him, and asked for and received his consent to search his luggage.[260] Justice O'Connor's majority opinion explicitly notes the police officers did not have even "articulable suspicion" that Bostick was a drug dealer, much less probable cause.[261] She states "even when officers have *no basis* for suspecting a particular individual, they may . . . request consent to search his or her luggage, . . . as long as the police do not convey a message that compliance with their request is required."[262] So requests for consent do fall under the rubric "it never hurts to ask."[263]

---

256. *Hicks*, 480 U.S. at 327.

257. It is reasonable to assume it is the police who initiate a request for consent. It is possible a person might walk up to an officer and offer to consent to search, but it is hardly likely that this is the typical scenario.

258. 111 S. Ct. 2382 (1991). *See also* Florida v. Royer, 460 U.S. 491 (1983).

259. 111 S. Ct. at 2384.

260. Actually Bostick consented to the search of two bags, the second of which was not his. If any contraband had been found in the second bag, that search could have been defended only under *Rodriguez*. *See* Kamisar, *supra* note 194, at 6 n.1.

261. *Bostick*, 111 S. Ct. at 2384, 2385-86. Justice O'Connor quotes the Florida Supreme Court's statement that the officers acted "admittedly without articulable suspicion." *Id.* at 2384-85 (quoting Bostick v. State, 554 So. 2d 1153, 1154-55 (1989)).

262. 111 S. Ct. at 2386 (citing Florida v. Royer, 460 U.S. 491, 501 (1983)) (emphasis added).

263. The analysis in Justice Scalia's majority opinion in California v. Hodari D., 111 S. Ct. 1547 (1991), seems to adopt a rule that it never hurts for a police officer to leap. In *Hodari*, a police officer who did not even have reasonable suspicion, much less probable cause, chased one of several youths who ran when the police approached. The youth allegedly tossed a rock of crack cocaine away "a moment" before the officer tackled him. *Id.* The Court ruled, seven to two, that the officer's lack of cause to chase the youth was immaterial because no seizure occurred until the officer actually grabbed the youth. Therefore, the youth's aban-

Justice O'Connor's explanation for this position is revealing. She justifies the absence of any scrutiny of the reasonableness of the police request for consent by stating the request occurred within "'the sort of consensual encounter[s] *that implicat[e] no Fourth Amendment interest.*'"[264] Justice O'Connor's conclusion that it never hurts for police to ask for consent rests, therefore, on the waiver conception of consent—the very conception of consent that Justice Scalia's opinion refuses to acknowledge.[265] Justice O'Connor, in other words, denies that consent is subject to any reasonableness criterion.

The inconsistent treatment of consent in *Bostick* and *Rodriguez* reveals that the rationale for "seemingly consented searches" in *Rodriguez* rests on an opportunistic picking-and-choosing of implications from the two inconsistent versions of consent. When *Rodriguez* declines to require the police to have any understandable reason (let alone probable cause) for requesting consent, it necessarily draws on the traditional conception of consent as a waiver or foregoing of the protection provided by the reasonableness standard (as Justice O'Connor does in *Bostick*), so that the reasonableness requirement does not apply to the police request. When the *Rodriguez* majority concludes a mere appearance of authority to consent is a sufficient basis for upholding an intrusion, however, it drops the waiver concept of consent (that can only be satisfied by valid authority) and switches to the claim that consent is an "element" that satisfies Fourth Amendment reasonableness. The majority still refuses, however, to acknowledge the probable cause criterion that is the essence of Fourth Amendment reasonableness.

This opportunistic have-your-cake-and-eat-it-too treatment of consent and reasonableness in *Rodriguez* is indefensible. Once the majority decided to reject the waiver concept of consent and to treat consent as an element that satisfies Fourth Amendment reasonable-

---

donment of the cocaine prior to any seizure did not implicate any Fourth Amendment interest. As I understand *Hodari*, the Court's view is that an officer does not have to have cause when he leaps at a citizen, although the Court does not dispute that the officer is required to have cause the instant he lands on the citizen. It seems to me *Hodari* gives insufficient weight to the principle (of physics, not law) that whoever leaps must come down.

264. 111 S. Ct. at 2386 (quoting Florida v. Rodriguez, 469 U.S. 1, 5-6 (1984) (emphasis added)). Justice O'Connor does not claim it was "reasonable" for the police to ask to search the luggage of a person against whom they had no articulable suspicion. *Bostick*'s consensual encounter rationale allows the police to act in ways that could not be justified under Fourth Amendment reasonableness. Contrast Terry v. Ohio, 392 U.S. 1 (1968), in which the Court stated "a police officer may in appropriate circumstances [i.e., when there is reasonable suspicion] and in an appropriate manner approach a person for purposes of investigating possibly criminal behavior even though there is no probable cause to make an arrest." *Id.* at 22.

265. *See* 110 S. Ct. at 2800 n.*.

ness, there was no longer any possible justification for failing to assess whether there was adequate cause for the police request for consent.[266] If consent and "seeming consent" are to be treated as providing a justification that satisfies Fourth Amendment reasonableness, then the cause requirement (and the probabable cause standard) should apply to the "element" of consent.[267]

The full implications of the *Rodriguez* majority's failing to inquire why the police want consent for an intrusion are obscured by the facts in *Rodriguez*. The officers who entered Rodriguez's apartment did have a substantial reason for wanting to enter; they wanted to arrest Rodriguez for beating Fischer. There is every reason to think a magistrate would have found the police had probable cause for an arrest warrant, if they had bothered to seek one.[268] The crucial point, however, is that the fact that the police had information amounting to probable cause is *irrelevant* to the holding announced in Justice Scalia's opinion.

Rodriguez does not make probable cause a condition for either a consented or "seemingly consented" intrusion.[269] To the contrary, Justice Scalia's opinion never actually states that the officers possessed (or even probably possessed) probable cause,[270] perhaps to avoid any suggestion that this could have mattered in the resolution of the case.

---

266. Justice Scalia claims consent operates as an exception to the warrant requirement. *See supra* notes 135-40 and accompanying text. The conclusion that consent is an exception to the warrant requirement, however, does not provide any basis for concluding the probable cause requirement does not apply. The exceptions to the warrant requirement that could apply to entry of a home do require probable cause. *See supra* notes 224-25 and accompanying text.

267. Both the damage *Rodriguez* does to constitutional doctrine and the threat it poses to citizens' privacy would have been substantially mitigated, though not entirely cured, if the Court had indicated that "seemingly consented" intrusions would only be deemed reasonable under the Fourth Amendment if the police possessed probable cause for the intrusion. Under this approach, which would have treated "seeming consent" as though it were an exception to the warrant requirement (the other enumerated exceptions to the warrant requirement that can apply to intrusions of residences require probable cause), much of the arbitrariness presently evident in Justice Scalia's treatment would have been eliminated. *See supra* notes 240-42 and accompanying text.

The fact that this approach was not taken, however, demonstrates that the result the majority sought to achieve in *Rodriguez* was to provide police with a way to justify intrusions for which probable cause could not be shown. If *Rodriguez* is considered in combination with *Schneckloth* and *Bostick*, it is obvious the majority Justices have undertaken to facilitate police claims of "consent" in order to allow intrusions in situations in which the police have not or cannot satisfy the probable cause requirement.

268. *See supra* note 44 and accompanying text.

269. Justice Scalia asserts the probable cause requirement does not apply to "consented" searches. 110 S. Ct. 2799.

270. This is noteworthy because respondent Rodriguez conceded probable cause existed for an arrest warrant. Respondent's Brief at 26.

The full scope of the power *Rodriguez* confers on the police through its endorsement of "seemingly consented" intrusions as "reasonable" intrusions emerges only if one looks beyond the specific entry of Rodriguez's apartment. Consider a hypothetical scenario. Assume the police decide to search every house in a neighborhood for contraband of any kind. Officers go to the door of one house and say to the babysitter who answers the door (she is not a resident) "We are asking all of the residents of this neighborhood to cooperate with our campaign to stamp out crime for once and for all by allowing us to search their homes for evidence of crime. Law-abiding citizens have nothing to fear. Will you cooperate with us and allow us to examine your home? Or do you have something to hide?" The babysitter, who is frightened of police generally, says "okay" and fails to point out that she is only the babysitter. The police enter, thoroughly search the house, and discover evidence of a crime in the closet of the master bedroom. (It does not matter what crime, there is no cause for the search.) Is this search of a home "reasonable" and constitutional?

The Rehnquist Court would probably answer yes. Under *Bostick* it was permissible for the police to request consent without cause. Under *Rodriguez* the Court would probably conclude there was "seeming consent" to the search of the house even though there was no valid consent. The police had some indications suggesting the woman was a resident, and this is probably all *Rodriguez* requires for "seeming consent."[271] There is, however, *neither cause for this search nor consent to it.* Any claim that such a search is "reasonable" is concocted entirely from a police error that resulted from an *unjustifiable* request.

Is it plausible the Framers intended Fourth Amendment *reasonableness* to be as puny as it is in this example? Is this the full extent to which the Framers sought to protect "the right of the people to be secure in their . . . houses"? Is this all they had in mind when they forbade "unreasonable searches"? How can it be "reasonable" for a citizen's privacy to be breached merely "because" a police officer was confused about who lived where, for whatever reason, if the officer cannot articulate any individualized cause for wanting to search in the first place?[272]

---

271. The police possessed some basis to warrant a belief the woman was a resident: (1) she answered the door, and (2) the police told her they were asking consent of the residents for a search of their houses and she did not indicate she was not a resident. This is probably enough to satisfy the *Rodriguez* objective standard for "seeming consent." *See supra* note 30.

272. Justice Scalia cites the following language from *Prouse*: "The essential purpose of the proscription in the Fourth Amendment is to impose a standard of 'reasonableness' upon the exercise of discretion by government officials." 110 S. Ct. at 2800 n.* (citing Delaware v. Prouse, 440 U.S. 648, 653-54 (1979)). Justice Scalia,

Rodriguez, calls for only a feeble assessment of the "reasonableness" of police conduct. It does not ask whether the police conduct was "reasonable" (even in the limited sense that it was "understandable") in all its facets. It only considers whether the officer possessed information that would "warrant" a belief that the consenting person possessed authority to give consent.[273] This information, however, can never suffice to provide cause for an intrusion. The *Rodriguez* majority is not candid when it speaks of "seeming consent" as "satisfying" the Fourth Amendment while omitting the probable cause requirement.[274]

D. Why Not Requiring Magistrate Approval for "Seemingly Consented Searches" Is Anomalous

Justice Scalia's assertion that the warrant requirement does not apply to "seemingly consented searches" is a necessary component of any rationale for upholding the police intrusion of Rodriguez's apartment.[275] This assertion is not surprising in view of the Court's

---

however, omits to say the standard of reasonableness to which *Prouse* refers is a *cause* standard.

In *Prouse*, Justice White presents the following description of the implications of the reasonableness standard:

> [T]he permissibility of a particular law enforcement practice is judged by balancing its intrusion on the individual's Fourth Amendment interests against its promotion of legitimate governmental interests. Implemented in this manner, the reasonableness standard usually requires, at a minimum, that the facts upon which an intrusion is based be capable of measurement against "an objective standard," whether this be probable cause or a less stringent test.

440 U.S. at 654 (footnotes deleted).

As Professor Kamisar has written, "a 'reasonable' suspicion*less* search is a Fourth Amendment oxymoron." Yale Kamisar, *Prepared Remarks at U.S. Law Week's Eleventh Annual Constitutional Law Conference* 15 (1989) (mimeo).

273. Whatever "warrant" means. *See supra* note 30.

274. The footnote in Justice Scalia's opinion demonstrates a studied avoidance of any consideration of the probable cause requirement in *Rodriguez*. Justice Scalia wrote "The only basis for contending that the constitutional standard [of reasonableness] could not possibly have been met [in a seemingly consented search] is the argument that reasonableness must be judged by the facts as they were, rather than by the facts as they were known." 110 S. Ct. at 2800, n.*. That is incorrect. The search was unreasonable because it did not satisfy the probable cause and warrant requirements that constitute the standard of Fourth Amendment reasonableness.

275. If the warrant requirement were applied to the police entry of Rodriguez's apartment, the result would have been unambiguous. As Justice Marshall wrote, "the warrantless entry of Rodriguez's apartment must be viewed as an unreasonable search in violation of the warrant requirement of the Fourth Amendment and the subsidiary rule that 'searches and seizures inside a home without a warrant are presumptively unreasonable.'" 110 S. Ct. at 2802 (Marshall, J., dissenting) (quoting Payton v. New York, 445 U.S. 573, 586 (1980)).

reluctance to favor warrant intrusions over consented intrusions.[276] Even so, the nonapplication of the warrant requirement creates several noteable anomalies.[277]

The most glaring anomaly arises from the amount of discretion that the police are accorded in deciding whether to enter a home. Under *Rodriguez*, the police have discretion to decide for themselves prior to an intrusion whether a third party possesses the requisite authority to consent. *Payton* held, however, the police do not have similar discretion to assess probable cause for a similar intrusion.[278] Why should police officers be allowed to enter a home on the basis of *their own assessment* of a third-party's status and authority when they are not allowed to enter on the basis of their own assessment of probable cause?

It cannot seriously be contended that the police are inherently better equipped to assess third-party authority than they are to assess probable cause. The police error in *Rodriguez* demonstrates an assessment of a third-party's authority to consent is as vulnerable to police overzealousness as is an assessment of probable cause. The

---

276. There is language in then-Justice Rehnquist's majority opinion in *Gates* that appears to indicate a preference for warrant searches over consent searches. Specifically, Justice Rehnquist inveighed against imposing an overly exacting probable cause standard for warrants because "the police might well resort to warrantless searches, with the hope of relying on consent." *Gates*, 462 U.S. at 236. It appears, however, that this statement was inserted merely to make the relaxation of the probable cause standard in *Gates* more palatable. Justice Stewart's earlier opinion in *Schneckloth* clearly did not express any preference for warrant searches over consent searches. *See Schneckloth*, 412 U.S. at 227-28.

Justice Marshall, however, expressed a preference for warrant searches over consent searches in his *Rodriguez* dissent:

> third-party consent searches are not based on an exigency and therefore serve no compelling social goal. Police officers, when faced with the choice of relying on consent by a third party or securing a warrant, should secure a warrant, and must therefore accept the risk of error should they instead choose to rely on consent.

110 S. Ct. at 2804.

277. It may appear there is an obvious practical reason why the police should be able to act without a warrant in the instance of a consented (or "seemingly consented") intrusion: the police are on the scene but the magistrate is not. Gaining the magistrate's approval will be time-consuming and inconvenient. It is not inconceivable that the "consenting party" might sometimes change his or her mind in the interim. For example, the delay might allow the third party to communicate either with the suspect or with an attorney who might dissuade the third party from further cooperation with the police. The possibility that this might occur, however, should not constitute a legitimate ground for framing the requirements for a supposedly voluntary act of cooperation. *Cf. Schneckloth*, 418 U.S. at 248 (consent is ineffective if given as a result of duress or coercion, express or implied).

Inconvenience, moreover, is not an adequate reason to excuse the warrant requirement; it is always inconvenient to get a warrant. *See* Johnson v. United States, 333 U.S. 10, 15 (1948).

278. *See supra* note 51 and accompanying text.

purported police error in *Rodriguez* might well have been avoided if the officers had been required to demonstrate Fischer's authority to consent to a magistrate.[279] Recall that Justice Scalia agreed the record of the suppression hearing demonstrated it was clear and obviously correct that there was no showing that Fischer possessed authority under *Matlock*.[280]

The omission of the warrant requirement from the purported element of "seeming consent" also creates an anomaly in terms of the discretion the police have as to the manner and extent of their intrusion into a residence. For example, if the police officers who entered Rodriguez's apartment had obtained an arrest warrant for battery, they probably would not have been allowed to enter Rodriguez's apartment without knocking or alerting him to their presence (as they did) because there was nothing in the situation that would have justified the authorization of a "no-knock" warrant.[281] The police, however, were able to enter and surprise a sleeping suspect by acting under Fischer's "seeming consent." It is even likely the officers chose to act on her "consent" to their entry of Rodriguez's apartment because of this opportunity for surreptitious entry.[282]

Third-party "seeming consent" also may provide police with a broader license to roam through a residence than an arrest warrant

---

279. I concede I am glossing over the problem of magistrates rubber-stamping police requests for warrants. The existing empirical research suggests this is a significant limitation on the value of the warrant process. *See* VAN DUIZEND, *supra* note 17, at xi-xii, 47-49, 95. Even so, it is likely the warrant process often exerts a beneficial influence in terms of protecting citizens' privacy. The process of preparing a warrant application almost certainly prompts discussion between police officers and supervisors. This discussion might serve to eliminate some groundless intrusions. In addition, in many urban jurisdictions prosecutors (whose office will be called upon to defend the legality of the resulting intrusions) screen warrant applications before they are submitted to magistrates. The empirical literature suggests this review is often more substantial than that provided by magistrates. *See* VAN DUIZEND *supra* note 17, at 82-98. Thus, the warrant process is beneficial in inhibiting arbitrary police conduct notwithstanding that it does not operate consistently with the formal model.

280. *See supra* text accompanying notes 67-68. The police could have obtained all of the information appearing in the record of the suppression hearing by asking Fischer additional questions.

281. Statutes in a number of jurisdictions restrict the manner in which warrants may be executed by requiring the police to knock and make their presence known to the inhabitants of a dwelling prior to any forcible entry. The Court has indicated, however, that these restrictions on the execution of warrants are not required by the Constitution. *See* Dalia v. United States, 441 U.S. 238, 257 n.19 (1979).

282. Obviously the officers also would have lost the element of surprise if they had simply knocked on Rodriguez's door and sought his own consent to their entry. Moreover, if they proceeded in this fashion they ran the risk that he might not be home, thereby eliminating any justification for entering his apartment at all. By relying on Fischer's "consent," however, they could enter and inspect Rodriguez's apartment even if he were absent.

It is clear the police suspected there were drugs in Rodriguez's apartment. *See supra* text accompanying note 33.

would. Unless the consenting party expressly limits the scope of consent to enter a residence, the police typically will be empowered to look through the entire premises,[283] which provides great latitude for "in plain view" discoveries. In contrast, the scope of a police examination of a residence allowed by an arrest warrant may be more limited. An arrest warrant empowers the police to intrude into a residence only to the extent needed to make the arrest.[284] For example, if the police arrest a suspect in the living room of an apartment, they would not have any justification for entering a bedroom.[285] Hence, an intrusion pursuant to an arrest warrant often would not bestow the same unlimited license to roam through a residence that third-party "seeming consent" would.[286]

---

283. Suppose a police officer requests permission to look around a house and a person who appears to be a resident consents. This would constitute consent for a search of the entire house, would it not? *See generally* Florida v. Jimeno, 111 S. Ct. 1801, 1803-04 (1991) (defendant's general consent to search of his car for drugs, absent explicit limitation, conveys permission to search the entire car and the contents of any container found within it).

284. *See* Maryland v. Buie, 110 S. Ct. 1093 (1990). Justice White's majority opinion states "[U]ntil the point of Buie's arrest the police had the right, based on the authority of the arrest warrant, to search anywhere in the house that Buie might have been found." *Id.* at 1096. He also notes that after the suspect is found, "the search for him was over, and there was no longer that particular justification for entering any rooms that had not yet been searched." *Id.* at 1097.

285. In that situation, the police could make an intensive search of the living room under the search incident to arrest doctrine. A search incident to arrest is intended to prevent the arrestee from gaining access to weapons and from destroying evidence. For this reason, it is limited to the "area within the immediate control of the arrestee." Chimel v. California, 395 U.S. 752, 763 (1969). The area of immediate control is usually interpreted to mean the room in which the arrest occurs, though it might also include other rooms through which the arrestee must be taken to exit the residence. The police also may seize evidence in another room that could be observed in plain view from the living room. They would not have any other authority to enter or examine other rooms, however.

The Court recently approved a "safety sweep" of a residence in Maryland v. Buie, 110 S. Ct. 1093 (1990). This sweep occurred in a context in which the police had reason to believe that the robbery suspect being arrested had an unapprehended accomplice. *Id.* at 1093. In the majority opinion, Justice White held "that there must be articulable facts which, taken together with the rational inferences from those facts, would warrant a reasonably prudent officer in believing that the area to be swept harbors an individual posing a danger to those on the arrest scene." *Id.* at 1098. It remains to be seen how expansive this safety sweep doctrine will become.

286. A search pursuant to a search warrant is expressly limited by the statement in the warrant of the particular place to be searched and the particular items for which the police are searching. Thus, in theory, a search warrant limits the scope of the search that it authorizes. *See Rodrgieuz*, 110 S. Ct. at 2806 (Marshall, J., dissenting).

The reality may be otherwise. An empirical study of the warrant process reports many warrants include items like rent receipts for the purpose of expanding the scope of the authorized search. Thus, search warrants may effectively authorize the search of an entire residence in practice. *See* VAN DUIZAND, *supra* note 17, at 29.

These anomalies have at least two significant implications. First, they indicate a resident's privacy in his or her home is significantly more vulnerable to intrusions based on "seeming consent" than to intrusions based on the more rigorous review entailed in the warrant process. This discrepancy is indefensible; highly invasive intrusions should be bounded by the nature of the cause for the intrusion. Second, one can expect these anomalies will encourage police officers to seek out third-party "seeming consent" instead of obtaining a warrant. The anomalies are such that police officers probably will see obtaining "seeming consent" as a way to maximize their discretion to act as they think best.[287] *Rodriguez*, in short, reflects the current majority's disinterest in the warrant requirement—the procedure the Fourth Amendment mandates to protect the privacy of citizens' homes from exactly the kind of overbearing police conduct that occurred in *Rodriguez*.

In summary, the treatment of Fourth Amendment reasonableness in *Rodriguez* reflects the Rehnquist Court's disregard for traditional Fourth Amendment principles and doctrine. The majority's rationale in *Rodriguez* depends on the substitution of a colloquialized notion of "reasonableness" for the traditional concept of Fourth Amendment reasonableness. This substitution, however, trivializes the meaning of Fourth Amendment reasonableness when it is applied in the context of a highly invasive police intrusion of a home. In making this substitution, *Rodriguez* cuts the "reasonable" label "loose from its theoretical and practical moorings."[288]

If the *Rodriguez* majority had wanted to claim that police intrusions of homes based only on the "seeming consent" of a third party satisfy Fourth Amendment reasonableness, then they should at least have incorporated a probable cause requirement into the new "seeming consent" warrant exception. *Rodriguez*'s claim that a highly invasive warrantless intrusion of a home can be "reasonable" without any assessment of cause is a distortion of the core idea of the Fourth Amendment.

VI. HOW *RODRIGUEZ* OVERSTATES THE EXCUSABILITY OF UNDERSTANDABLE POLICE ERRORS

The third salient claim in the rationale for *Rodriguez* is that understandable police mistakes about facts do not contravene the

---

287. This is especially true given the minimal showing of voluntariness required by *Schneckloth* and the opportunity for police to restructure who said what to whom and to filter what "information [was] available to [them] at the moment." *See infra* note 311.

In the past, there was one constraint on police officers' reliance on dubious third-party consent claims. They knew the fruits of their intrusion would be lost under *Matlock* if it turned out the third party lacked authority to consent. This constraint effectively disappears with the *Rodriguez* majority's approval of "seeming consent."

288. Arizona v. Hicks, 480 U.S. 321, 326 (1987).

reasonableness requirement of the Fourth Amendment. Justice Scalia asserts the Fourth Amendment does not require factual accuracy.[289] He supports this claim by offering several examples of police errors about facts that do not violate the Fourth Amendment. He notes a warrant is not invalid simply because it is issued on the basis of "factually inaccurate information."[290] Similarly, he states police officers do not violate the Fourth Amendment if they enter a residence without a warrant "because they reasonably (though erroneously) believe that they are in pursuit of a violent felon who is about to escape."[291] He also cites three cases as authority for a "general rule"[292] that police errors about "factual determinations bearing upon search and seizure"[293] do not violate the Fourth Amendment if they are "reasonable" (understandable) errors: *Brinegar v. United States*,[294] *Hill v. California*,[295] and *Maryland v. Garrison*.[296]

There are two defects in this claim. First, the characterization of the police conduct in *Rodriguez* as a "factual error" is arbitrary.

---

289. 110 S. Ct. at 2799.

290. *Id.* at 2799. Justice Scalia wrote:

> If a magistrate, based upon seemingly reliable but factually inaccurate information, issues a warrant for the search of a house in which the sought-after felon is not present, has never been present, and was never likely to have been present, the owner of that house . . . does not suffer a violation of the Fourth Amendment.

Id. Note that this warrant search example is not comparable to a "seemingly consented search." If there is no violation of the Fourth Amendment in this example, it can only be because there is probable cause (albeit constructed from inaccurate information) for a valid warrant. Inaccurate information does not preclude a warrant from being valid. *See* Franks v. Delaware, 438 U.S. 154 (1978) (warrant based on inaccurate information presented in police affidavit is valid unless the officers recklessly or deliberately misled the magistrate). *Rodriguez*, however, does not require any assessment of probable cause in a "seemingly consented search." *See supra* text accompanying notes 269-70.

291. 110 S. Ct. at 2800 (citing Archibald v. Mosel, 677 F.2d 5 (1st Cir. 1982)). Justice Scalia wrote that the Fourth Amendment

> is no more violated when officers enter without a warrant because they reasonably (though erroneously) believe that the person who has consented to their entry is a resident of the premises than it is violated when they enter without a warrant because they reasonably (though erroneously) believe they are in pursuit of a violent felon who is about to escape.

Id. These two situations are *not* comparable, however: the pursuit example involves a constitutional search only if the erroneous information possessed by the officers constitutes probable cause for the "hot-pursuit" entry (that is, the same showing of probable cause required for issuance of a search warrant for the wanted felon). *See* Warden v. Hayden, 387 U.S. 294, 298-99 (1967). *Rodriguez*, however, does not require any assessment of probable cause in a "seemingly consented" entry. *See supra* text accompanying notes 269-70.

292. *Rodriguez*, 110 S. Ct. at 2800.

293. *Id.* at 2801.

294. 338 U.S. 160 (1949).

295. 401 U.S. 797 (1971).

296. 480 U.S. 79 (1987).

Even assuming the police conduct was the result of a genuine mistake and not a willful disregard of Rodriguez's right, the mistake could have been the product of police ignorance of the *Matlock* co-inhabitant standard for authority to consent rather than the product of a factual misunderstanding of Fischer's status. In fact, the record in *Rodriguez* indicates that the police probably were ignorant of the *Matlock* co-inhabitant standard. This likelihood is significant because it is hard to see how one could treat police ignorance of search law standards as "reasonable" or "understandable."

Second, assuming *arguendo* the police did make a factual error about Fischer's residence, the general rule alluded to in *Rodriguez* still overstates the excusability of police errors. The examples and cases that Justice Scalia cites do not show that *any* "factual" error made by the police is tolerable under the Fourth Amendment if it is understandable in the circumstances. Rather, they stand for a more precise and qualitatively different proposition: *understandable police errors do not violate the Fourth Amendment if they involve a fact (or facts) that runs to the assessment of probable cause (or exigency) for an intrusion.* Allowing an understandable error about a fact that contributes to a showing of probable cause is not controversial because probable cause is a probabilistic standard by definition. Therefore, such an error does not contravene any legal standard.

The alleged police error in *Rodriguez*, however, is unrelated to probable cause. As discussed in the preceeding part, no assessment of cause is involved in a "seemingly consented" intrusion. The purported police error in *Rodriguez* regarding Fischer's authority to consent relates to whether there was legal authority for the intrusion.[297] An error about legal authority for an intrusion does breach a legal standard, and no precedent supports the conclusion that a police error that relates to the validity of legal authority can satisfy the Fourth Amendment. Indeed, the inexcusability of errors regarding legal authority is the premise for the Court's creation of exceptions to the exclusionary rule in both *United States v. Leon*[298] and *Illinois v. Krull*.[299]

Understandable police errors about facts that run to assessments of probable cause, including assessments of exigent circumstances, are properly excusable under the Fourth Amendment. However, other

---

297. Some source of legal authority for police conduct must exist. *See supra* note 190. Warrants can provide such authority. If the intrusion can be justified under an exception to the warrant requirement, the authority conferred on police officers by statute will suffice. If there is neither a warrant nor an applicable exception to the warrant requirement, the only possible source of authority for the police conduct is the consent of a resident who possesses authority to allow the police to enter a residence.

298. 468 U.S. 897 (1984).

299. 480 U.S. 340 (1987).

kinds of police errors, regardless of whether they can be labeled factual or not, should not be excused.

A. *Rodriguez*'s Arbitrary Assignment of the "Factual" Label to Police Errors about Authority to Consent

Justice Scalia's rationale in *Rodriguez* is constructed around an assumption that the police misassessment of Fischer's authority to consent was the product of a factual misunderstanding by the officers of Fischer's place of residence. This assumption is critical from a doctrinal perspective because precedent exists for excusing some types of factual errors made by the police, but no doctrinal basis exists for excusing police ignorance of legal standards. The assumption that the police misassessment of Fischer's authority to consent was factual in nature, however, is also arbitrary.

The question the police were required to answer in assessing Fischer's authority to consent did not consist only of a "factual determination" about where she resided. In Justice Scalia's words, the question was whether "the consenting party had authority over the premises."[300] That question does not simply implicate facts; it calls for a legal conclusion that involves a compound judgment of facts and law. The police cannot make a credible assessment of "authority over the premises" unless they start with a correct understanding of the *Matlock* co-inhabitant legal standard for authority to consent.[301] A police error about authority to consent could be the result of a misunderstanding about where the "consenting" party lived. It just as readily could be the result of police ignorance of the co-inhabitant legal standard.[302]

The possibility of police ignorance of the legal standard is demonstrated by the record in *Rodriguez*. Although no factual finding was made about the nature of the purported police error (or even as

---

300. *See supra* text accompanying note 83.

301. 110 S. Ct. at 2800. It is true that the *Matlock* co-inhabitant standard for common authority implicates facts, including the living arrangement of the third party and the acceptance of the other residents, but this hardly makes the issue of a third party's legal authority only a factual determination. A conclusion regarding Fischer's authority to consent requires two premises: 1) the legal standard to be applied (i.e., *Matlock*), and 2) the facts relevant to the application of the standard. Justice Scalia's "factual" label collapses this two-step assessment and arbitrarily omits any inquiry into whether the police applied the proper legal standard.

302. The shift from an actual authority standard to a "seeming consent" standard has the effect of making police knowledge of the law more problematic. If a court applied the *Matlock* standard, police knowledge of the co-inhabitant standard was irrelevant because the focus of inquiry was entirely on the actual living arrangements of the consenting party. Under a "seeming consent" standard, the inquiry is directed to the police assessment of authority, which should involve the question of whether the police knew and applied the correct legal standard.

to whether there was any genuine police error at all), two aspects of the police testimony in the record provide circumstantial indications that the police officers who entered Rodriguez's apartment did not comprehend the co-inhabitant legal standard.[303]

The first aspect of the police testimony that casts doubt on the officers' knowledge of the co-inhabitant standard is Officer Entress's testimony at the preliminary hearing. At that hearing he testified Fischer told him she "used to live" in the apartment.[304] If this testimony is accurate then the officers knew Fischer was not a co-inhabitant on the date of the entry and therefore clearly could not have relied on her "consent" unless they were ignorant of the *Matlock* standard.

Officer Entress subsequently changed his testimony and said, a full year after the entry, that Fischer actually told him she "had been living" at the apartment.[305] Putting aside the obvious possibility of perjury, the change in Entress's testimony does not disturb the implication that, at a minimum, he was unfamiliar with the *Matlock* standard when he first testified, and by implication, when the entry occured. It is simply not plausible that an officer who understood the co-inhabitant standard would have been so careless in testifying that he would misstate such an important piece of information relating to Fischer's authority to consent.[306]

A second aspect of Entress's testimony also suggests the police officers were ignorant of the co-inhabitant standard when they acted on Fischer's "consent." He testified Fischer told the officers she had possessions and clothes in the apartment they wanted to enter.[307] This is a rather peculiar topic for discussion if the police actually thought that Fischer lived in the apartment. People normally have their possessions and clothes where they reside.

The fact that the topic was discussed at all calls for an explanation. The most plausible explanation is that the topic came up because the police were aware Fischer had moved out, but wrongly believed the

---

303. No testimony in the record directly explores the officers' knowledge of the *Matlock* standard for third-party consent at the time of the search.

304. *See supra* notes 34-35 and accompanying text. As indicated, Justice Scalia's opinion finesses the significance of the inconsistent police testimony.

305. *See supra* note 36 and accompanying text.

306. What Fischer told the police is relevant evidence on the issue of whether she resided in the apartment in fact, not merely as to what the police could believe. Even the statement Officer Entress attributed to Fischer in his revised testimony—she "had been living" in the apartment—is ambiguous, however, as to whether Fischer was living in the apartment at the time of the intrusion. *See supra* note 36 and accompanying text. Yet Entress testified he did not attempt to clarify her answer. Could an officer who understood *Matlock* have accepted the ambiguous statement attributed to Fischer?

307. *See supra* note 42. The police testimony was that Fischer included "all of her clothing" among her possessions in the apartment.

fact that she still had possessions in the apartment was enough to give her authority to consent. These two aspects of the police testimony in *Rodriguez* suggest the police error was a legal error; it appears the officers were ignorant of *Matlock*'s co-inhabitant standard.[308]

It *should* matter whether the officers were confused about the facts or the law. There is no precedent for excusing police ignorance of the law because police officers are obligated to know the law.[309] There is no jurisprudential basis for labeling a police error about a legal standard "reasonable" police conduct.[310] This, presumably, is why the rationale for *Rodriguez* strains to treat the police mistake as "factual" in nature. Rhetorical necessity is not, however, an adequate excuse for arbitrary characterization. One of the most unfortunate practical consequences of the "seeming consent" approach adopted in *Rodriguez* is the opportunity it creates for the transformation of police ignorance of the law into "understandable" errors "about facts" through the restructuring of police accounts of the circumstances leading up to a "seemingly consented search."[311] The willing-

---

308. It is important to be clear about the belief that has to be warranted for "seeming consent." It is not enough that the police had information that suggested in a general way that Fischer could give consent. *Matlock* sets a co-inhabitant standard that required the police to have information to support a reasonable belief that Fischer was a co-inhabitant.

309. *See infra* text accompanying notes 363-69.

310. This is the fundamental flaw in the proposal for a good-faith mistake exception to the exclusionary rule that would be broad enough to cover police legal errors in warrantless searches. What criteria could be used to determine which police legal errors are "reasonable"? Justice White proposed that police errors should be excused if "an officer act[ed] in the good-faith belief that his conduct comported with existing law and [had] reasonable grounds for this belief." Stone v. Powell, 428 U.S. 465, 538 (1976) (White, J., dissenting). How, however, can a police error about a legal standard possibly be based on existing law?

311. One of the strengths of the *Matlock* actual authority standard was that it was susceptible to corroborable evidence regarding the residence of the "consenting" third party. Under a "seeming consent" standard, however, the relevant evidence will consist largely of noncorroborable statements regarding what the third party told the police. In that setting, evidence given by the third party may be of dubious credibility; a third party who gave consent may well regret having done so.

The potential for police perjury is also present. One discussion of police ethics characterizes police perjury as so widespread that it is one of eight basic orientations that police recruits learn: "Lying and deception are an essential part of the police job, and even perjury should be used if it is necessary to protect yourself to get a conviction on a 'bad guy.'" Lawrence Sherman, *Learning Police Ethics*, 1 *Crim. Just. Ethics* 10, 15 (1982). *See also* People v. McMurty, 314 N.Y.S.2d 194 (N.Y. Crim. Ct. 1970); P. CHEVIGNY, POLICE POWER 187-88 (1969); LYNN M. MATHER, PLEA BARGAINING OR TRIAL? THE PROCESS OF CRIMINAL-CASE DISPOSITION 27, 72-76, 98 (1979); JONATHON RUBENSTEIN, CITY POLICE 386-90 (1973); SKOLNICK, *supra* note 9, at 215; RICHARD S. UVILLER, TEMPERED ZEAL: A COLUMBIA LAW PROFESSOR'S YEAR ON THE STREETS WITH THE NEW YORK CITY POLICE DEPARTMENT 111-18 (1988);

ness of the majority in *Rodriguez* to overlook the possibility of police ignorance of *Matlock* and to assume police misassessments of authority to consent are "factual" is indicative of the majority's predisposition to defend aggressive police conduct.[312]

B. Rodriguez*'s Overstatement of the Excusability of Understandable Police Errors about Facts*

Even if one assumes the police intrusion of Rodriguez's apartment were produced by a factual police error about Fischer's residence,[313]

VAN DUIZEND, *supra* note 17, at 56, 78-79, 100; Sarah Barlow, *Patterns of Arrests for Misdemeanor Narcotics Possession: Manhattan Police Practices* 1960-62, 4 CRI. L. BULL. 549, 557 (1968); Joseph D. Grano, *A Dilemma for Defense Counsel: Spinelli-Harris Search Warrants and the Possibility of Police Perjury*, 1971 U. ILL. L. F. 405, 408-09; Myron W. Orfield, Comment, *The Exclusionary Rule and Deterrence: An Empirical Study of Chicago Narcotics Officers*, 54 *U. Chic. L. Rev.* 1016, 1049-51 (1987); Myron W. Orfield, *The Exclusionary Rule in Chicago Courts*, 63 Colo. L. Rev. (forthcoming 1991) (reporting that judges, public defenders and prosecutors perceive widespread police perjury in connection with motions to suppress); Comment, *Police Perjury in Narcotics "Dropsy" Cases: A New Credibility Gap*, 60 GEO. L. J. 507 (1971); Note, *Effect of* Mapp v. Ohio *on Police Search-and-Seizure Practices in Narcotics Cases*, 4 COLUM. J.L. & SOC. PROBS. 87 (1968).

Defense attorneys also have expressed concern about police perjury. *E.g.*, Fred Cohen, *Police Perjury: An Interview with Martin Garbus*, 8 CRIM. L. BULL. 363, 365 (1972); Charles M. Sevilla, *The Exclusionary Rule and Police Perjury*, 11 SAN DIEGO L. REV. 839 (1974).

Critics of the exclusionary rule have also noted widespread police perjury but have suggested this is one of the costs produced by the rule. *E.g.*, MALCOLM WILKEY, ENFORCING THE FOURTH AMENDMENT BY ALTERNATIVES TO THE EXCLUSIONARY RULE 36 (National Legal Center for the Public Interest, 1982); Edna F. Ball, *Good Faith and the Fourth Amendment: The "Reasonable" Exception to the Exclusionary Rule*, 69 J. CRIM. L. & CRIMINOLOGY 635, 655 (1978); James Duke Cameron & Richard Lustiger, *The Exclusionary Rule: A Cost-Benefit Analysis*, 101 F.R.D. 109, 137-38 (1984); Oaks, *supra* note 9, at 739-42; James E. Spiotto, *Search and Seizure: An Empirical Study of the Exclusionary Rule and Its Alternatives*, 2 J. LEGAL. STUD. 243, 275-76 (1973).

There are some who dispute the prevalence of police perjury, however. *See, e.g.*, McRae v. Illinois, 386 U.S. 300, 313 (1967) (court not required to assume police perjury); D. Lowell Jensen and Rosemary Hart, *The Good Faith Restatement of the Exclusionary Rule*, 73 J. CRIM. L. & CRIMINOLOGY 916, 935 (1982) (reference to police perjury in connection with searches is "completely unfounded" and amounts to an "unfair, gratuitous slur on police").

312. The *Rodriguez* opinion does not mention the possibility that the police acted out of ignorance of the co-inhabitant legal standard. Justice Scalia's statement of the "objective standard" to be applied by lower courts in assessing whether police misassessments of third-party authority were understandable does not state any requirement that those courts are to determine whether the officers were familiar with the *Matlock* standard. *See supra* note 30.

313. For purposes of the discussion in this part of the article, however, I put aside my doubts that the police actually made the error Justice Scalia assumes they did.

the general rule of the excusability of such errors that is alluded to in *Rodriguez* is overstated. Justice Scalia offers three precedents for the purported general rule: *Brinegar*, *Hill*, and *Garrison*. An examination, however, shows each case involved factual aspects of the assessment of probable cause. The discussion of police errors in both *Brinegar* and *Hill* are simply explications of the probable cause standard. The discussion of police errors in *Garrison* is also embedded in the assessment of probable cause for the search in that case. This aspect of *Garrison* is not as evident as it is in *Brinegar* and *Hill* because Justice Stevens's majority opinion in *Garrison* contains loose dicta that overrun the dimensions of the actual police mistake. In that sense, *Garrison* does foreshadow the overgeneralized approval of police errors that appears in *Rodriguez*, but it does not provide credible authority for the "general rule" invoked in *Rodriguez*. Consequently, the authorities Justice Scalia cites in *Rodriguez* do not support his claim.

1. *Brinegar* and *Hill* Only Explicate the Probablistic Nature of Probable Cause

Justice Scalia cites *Brinegar* as though it generally supported excusing understandable factual errors made by the police that relate to justifications for intrusions.[314] In fact, however, there was no police mistake in *Brinegar*.[315] *Brinegar* only explicates the nature of the probable cause standard; the language that Justice Scalia quotes appears in a discussion contrasting the probable cause standard with the proof beyond a reasonable doubt standard applicable at trial.[316]

---

314. 110 S. Ct. at 4895. Justice Scalia quotes the following language from *Brinegar*:

> Because many situations which confront officers in the course of executing their duties are more or less ambiguous, room must be allowed for some mistakes on their part. But the mistakes must be those of reasonable men, acting on facts leading sensibly to their conclusions of probability.

Id. (quoting Brinegar v. United States, 338 U.S. 160, 176 (1949)).

315. *Brinegar* involved a warrantless search of an automobile that the police had probable cause to believe contained illegal liquor. The Court had approved warrantless searches of automobiles on probable cause in *Carroll v. United States*, 267 U.S. 132 (1925). The police in *Brinegar* did not make any error about the facts that constituted probable cause to believe the auto in question was being used to run illegal liquor. To the contrary, the Court treated the police assessment of probable cause to believe that the defendant was running liquor as correct.

316. The *Brinegar* majority ruled the lower court suppressed evidence erroneously because the lower court misperceived the probable cause standard. When it assessed probable cause in a suppression hearing, the lower court inappropriately excluded certain testimony regarding the information that the police had that the defendant was running liquor on the grounds that it was hearsay. *See* 338 U.S. at 172-76. The discussion in *Brinegar* explicitly addresses the assessement of probable cause. Justice Rutledge's majority opinion classified the crucial question presented in

It is farfetched to suggest *Brinegar* excuses understandable police mistakes except in the specific context of probable cause.[317]

Justice Scalia also quotes a passage from *Hill* as though it generally supported excusing understandable police errors relating to justifications for searches.[318] *Hill*, however, like *Brinegar*, simply explicated the probable cause standard. Police officers, who had probable cause to arrest Hill as a robbery suspect, went to Hill's apartment and arrested a man they found there whom they believed to be Hill. They then made a search of Hill's apartment incident to the arrest[319] and found evidence that was used to incriminate Hill in the robbery. It turned out the person they arrested was Miller, not Hill. By chance, Miller, who was in Hill's apartment, matched the physical description that the police had of Hill.[320]

The constitutionality of the seizure of the evidence from Hill's apartment depended on the constitutionality of Miller's arrest. Hill argued Miller's arrest was unconstitutional because the police did not have probable cause to believe Miller had committed a crime. The Court disagreed. Writing for the majority, Justice White stated "[t]he police unquestionably had probable cause to arrest Hill," and it was understandable the officers had mistaken Miller for Hill.[321] Justice White concluded "in these circumstances the police were entitled to do what the law would have allowed them to do if Miller had in fact been Hill, that is search incident to arrest . . . ."[322] When Justice

---

the case as "whether there was probable cause." *Id.* at 164. The majority concluded there was. In the course of reaching that conclusion, Justice Rutledge wrote: "In dealing with probable cause, . . . we deal with probabilities . . . . The standard of proof [to be applied in assessing probable cause in suppression hearings] is accordingly correlative to what must be proved." *Id.* at 175.

317. It must be noted that Justice Stevens's majority opinion in *Garrison* previously misconstrued the *Brinegar* language that Justice Scalia quotes in *Rodriguez*. *See infra* note 355.

318. Justice Scalia quotes the following language from *Hill*: "[S]ufficient probability, not certainty, is the touchstone of reasonableness under the Fourth Amendment and . . . the officer's mistake was understandable and the arrest a reasonable response to the situation facing them at the time." *Rodriguez*, 110 S. Ct. at 2800 (quoting *Hill*, 401 U.S. at 804).

319. The police made a warrantless search of Hill's entire apartment for several hours incident to Miller's arrest. This was clearly in excess of that allowed by Chimel v. California, 395 U.S. 752 (1969). *Chimel* restricted searches incident to arrest to the area in the immediate vicinity of the arrestee. However, the date of the search in *Hill* preceeded the announcement of *Chimel*, and the Court declined to apply *Chimel* retroactively. 401 U.S. at 802. *See* Maryland v. Garrison, 480 U.S. at 79 (Blackmun, J., dissenting) (questioning *Hill*'s precedential value).

320. 401 U.S. at 799.

321. *Id.* at 802-03.

322. *Id.* at 804. Justice White also agreed with the California Supreme Court that "[w]hen the police have probable cause to arrest one party, and when they reasonably mistake a second party for the first party, then the arrest of the second party is a valid arrest." *Id.* at 802 (citing People v. Hill, 446 P.2d 521, 523 (Cal. 1968)).

White wrote in *Hill* "sufficient probablility, not certainty, is the touchstone of reasonableness under the Fourth Amendment,"[323] he was writing in the context of the probable cause standard.[324] Therefore, it is not surprising the Justices unanimously held the existence of probable cause for the arrest of Hill justified the arrest of Miller and the subsequent search of the Hill's apartment.[325] *Hill*'s statement about "sufficient probability" is only an articulation of the meaning of probable cause.[326]

Neither *Hill* nor *Brinegar* supports the overblown general rule that Justice Scalia asserts. Neither case provides any ground for excusing police errors that run to the legal authority for an intrusion.[327]

2. What *Garrison* Did and Did Not Decide About Police Mistakes

Justice Scalia also cites language from *Garrison* that suggests "[t]he validity [of the search] depends on whether the officer's [mistake] was objectively understandable and reasonable."[328] Unlike *Brinegar* and *Hill*, *Garrison* does discuss "reasonable" police errors in language as colloquial and formless as that found in *Rodriguez*.

---

323. 401 U.S. at 804.

324. It is entirely possible that Justice White intentionally wrote this passage as broadly as possible to make it available for quotation in cases like *Rodriguez*. Commentators have previously noted Justice White's fondness for fact-based inquiries. *E.g.*, Dworkin, *supra* note 218. Justice White has also emerged as the Court's leading advocate of the proposal for a broad reasonable mistake exception to the exclusionary rule. *See* INS v. Lopez-Mendoza, 468 U.S. 1032, 1052 (1984) (White, J., dissenting); Illinois v. Gates, 462 U.S. 213, 254-67 (1983) (White, J., concurring).

325. Justice Harlan's dissenting opinion in *Hill* takes issue only with the majority's refusal to apply *Chimel* retroactively. *See supra* note 319.

326. Justice Marshall wrote "*Hill* should be understood no less than *Brinegar* as simply a gloss on the meaning of 'probable cause.'" 110 S. Ct. at 2806 n.2 (Marshall, J., dissenting). The *Hill* majority does not say in so many words that there was probable cause to arrest Miller. Rather it says the police were entitled to arrest Miller because there was sufficient probablility that Miller was Hill, and the police did have probable cause to arrest Hill. The Court may have avoided saying directly that there was probable cause to arrest Miller in order to avoid the controversy, which later emerged in *Garrison*, over whether probable cause must relate to a specific person. *See infra* notes 336-37 and accompanying text.

327. The discussion in *Hill* is so closely confined to the probable cause assessment that the Court never addressed whether the police had any legal authority to make the warrantless entry of Hill's apartment that preceeded Miller's arrest despite the fact that issue would be relevant to any ultimate conclusion as to the legality of the police conduct. The Court did not hold that a warrantless entry of a residence for the purpose of making an arrest is unconstitutional in the absence of an exigency until Payton v. New York, 445 U.S. 573 (1980). Thus, *Hill* is silent on the issue that is most analogous to the issue decided in *Rodriguez*: whether there was legal authority for the police entry.

328. 110 S. Ct. at 2799 (quoting Maryland v. Garrison, 480 U.S. 79, 88).

Garrison, however, does not provide sound authority for *Rodriguez* because its discussion of police error amounts to unnecessary dicta that runs far beyond the character of the mistake the police actually made.

Garrison involved a warrant search. Police officers learned from an informant that McWebb was selling drugs from his third floor apartment in a multi-unit building. After purportedly making a "reasonable investigation,"[329] the police determined McWebb had the only apartment on the third floor. On the basis of that information, they obtained a warrant that authorized a search of McWebb and "the premises known as 2036 Park Avenue third floor apartment . . . ."[330] However, the police were mistaken. There were actually two third-floor apartments that shared a common entry. One was McWebb's, and the other was Garrison's. While executing the search warrant, the police entered and observed Garrison's apartment before they realized they were not in McWebb's apartment.[331] They left Garrison's apartment after they realized their error, but not before seeing and seizing drugs.

The issue was whether the police error about the number of third floor apartments affected the validity of the seizure of the drugs from Garrison's apartment. Justice Stevens, writing for a six Justice majority, upheld the search. Justice Blackmun wrote the dissent. (Interestingly, the two authors changed camps in the voting in *Rodriguez*.[332])

---

329. 480 U.S. at 81. There was a difference of opinion on this point, however. Justice Blackmun argued in dissent that the officers' mistake was not understandable because there were a number of steps that they could have taken, but failed to take, to verify the number of third floor apartments. In particular, Justice Blackmun argued there was no justification for the officers' failure to question Garrison about his reason for being on the premises. If they had asked, they would have learned about his apartment before they entered it. *See* 480 U.S. at 99-100.

330. 480 U.S. at 82 n.3, 92.

331. The police mistake in *Garrison* clearly falls outside of the exception to the exclusionary rule created earlier in *Leon*. The *Leon* exception is explicitly limited to evidence seized within the terms of a search authorized by a facially valid, but actually invalid, search warrant. 468 U.S. 897, 918 n.19 (1984). The situation in *Garrison*, however, is not susceptible to the *Leon* characterization.

332. The fact that Justices Stevens and Blackmun changed sides in *Rodriguez* may be explainable. First, it is possible they were assigned to write the opinions in *Garrison* because they were the least committed members of the respective voting blocs. Chief Justice Rehnquist and Justice Brennan may have assigned the opinions to enlarge the number of adherents to the majority and dissenting opinions respectively.

One can also speculate about the specific differences between *Garrison* and *Rodriguez* that may have led the two Justices to change camps. For example, the fact that the police obtained a warrant in *Garrison* but not in *Rodriguez* may explain Justice Stevens's seemingly inconsistent votes. *See infra* note 362. Similarly, the fact that the police did have probable cause to arrest Rodriguez for battery (although this is not part of the rationale for the decision) may have led Justice Blackmun to be

The key to understanding *Garrison* is to focus on the nature of the disagreement between the majority and the dissent. On the surface, the disagreement concerns the scope of the search authorized by the search warrant.[333] Justice Stevens states "a literal reading of [the] plain language" of the warrant indicates it authorized a search of the entire third floor of the building.[334] Justice Blackmun's dissent took a different view. He read the warrant as applying only to McWebb's apartment because the probable cause showing for the warrant included information only about McWebb.[335]

This disagreement over the scope of the search authorized by the warrant reflects an underlying disagreement about the nature of probable cause. The literal language of the warrant was not the only reason Justice Blackmun read the warrant as he did. Rather, his reading derived from his understanding that probable cause must be established with regard to the privacy interest of a particular person. He reasoned there could not be probable cause regarding Garrison's apartment if the police had information pertaining only to McWebb. Thus, because there was no probable cause for a search of Garrison's apartment, the search warrant should not be read to authorize a search of Garrison's apartment.[336] In other words, the dispute in *Garrison* is ultimately traceable to two different understandings of the purpose of the Fourth Amendment. Justice Stevens's majority position treats the amendment as though it were essentially a regulatory scheme to control the conduct and discretion of police officers. Justice Blackmun's dissenting opinion treats it as though it demarcates a citizen's personal right to be left alone.[337]

---

more tolerant of the police error in *Rodriguez* than he was of the error in *Garrison* (where the police did not have any information linking Garrison to wrongdoing). *See infra* notes 336-37.

333. Justice Stevens wrote "The matter on which there is a difference of opinion concerns the proper interpretation of the warrant." *Garrison*, 480 U.S. at 82.

334. *Id.*

335. 480 U.S. at 92-93. The search of Garrison's apartment was necessarily unconstitutional under Justice Blackmun's reading of the warrant because it was a warrantless search of a residence in contravention of the rule that warrantless searches of residences are *per se* unreasonable in the absence of exigent circumstances. *Id.*

336. Justice Blackmun objected to the intrusion into Garrison's apartment because, unlike the search of *Hill*'s apartment, it invaded the privacy of a person against whom the police had no cause. He said:

> It may make some sense to excuse a reasonable mistake by police that produces evidence against the intended target of an investigation or warrant if the officers had probable cause for arresting that individual or searching his residence. Similar reasoning does not apply with respect to one whom probable cause has not singled out and who is the victim of the officers' error.

Id. at 95.

337. Justice Blackmun did not develop his point about the importance of the

One might have expected the constitutionality of the search of Garrison's apartment to have been straightforward once Justice Stevens's majority opinion concluded there was probable cause (despite the error about the number of apartments) for a search of the *entire* third floor[338] and read the language in the warrant to authorize a search of the *entire* third floor. Under that view of probable cause

---

target of probable cause in terms of the purpose of the Fourth Amendment or the nature of the right announced in the Amendment. His position, however, appears to reflect the view that the Amendment gives citizens a personal right of privacy. If the Amendment creates a personal sphere of privacy against government intrusion (it does after all speak of a right), the pertinent inquiry is not whether government agents possessed probable cause to breach someone's privacy, but whether the government had probable cause to breach the privacy of *this* particular citizen. Hence the government's cause for acting should be evaluated from the perspective of the citizen whose privacy was breached, not from the viewpoint of the government agent.

If this understanding of the Fourth Amendment is adopted, then there was no reasonable cause for the intrusion of Garrison's apartment. If the protection expressed in the Fourth Amendment is truly personal in nature, it should be immaterial whether the police could have known prior to the search whether there were two apartments. The essential fact is that the government did not have probable cause regarding Garrison. Indeed, one would conclude the search warrant would have been invalid as to Garrison, regardless of its wording, because a warrant must rest on probable cause, and the police had no information linking Garrison or his apartment to drug dealing.

On the other hand, if the amendment is understood to be aimed primarily at regulating the conduct of government agents, as Justice Stevens' majority opinion appears to treat it, then Justice Blackmun's concern is irrelevant. Under the regulatory conception, the only question is whether the government agents possessed reasonable cause to take the action they did. This implies reasonable cause should be assessed from the perspective of government agents. The police did have probable cause and obtained a warrant authorizing them to conduct a search. They did nothing culpable. Therefore, the government intrusion of Garrison's apartment under the search warrant that appeared to apply to the entire third floor was reasonable from this regulatory perspective.

338. The police mistake about the number of apartments in *Garrison*, like the police error in *Hill*, involved a fact that went to the assessment of probable cause. "The place to be searched" is important to the validity of a warrant in two distinct ways. First, it is included in the warrant to define the scope of the search the police are authorized to make. A warrant that does not particularly state the place to be searched is invalid because it does not constrain the authority conferred on the police. The statement of the place to be searched is not probabilistic in its role as a limitation upon the scope of the authorized search.

"The place to be searched" is also an integral part of the probable cause showing for the warrant. Probable cause for a search requires more than information about the suspect's conduct; probable cause necessarily requires information about where the evidence of that conduct or contraband can be found. Probable cause for a search necessarily implicates both what the police expect to find and where they expect to find it. Thus, for the purposes of determining the validity of a warrant (as opposed to the scope of the authorized search), the statement of the place to be searched in the warrant should be assessed as a statement of probablility reflecting what the police can "reasonably know prior to their search." Therefore, the police error about the number of third floor apartments in *Garrison* ran to probable cause.

and that reading of the warrant language, the police discovery and seizure of the drugs during the execution of the valid warrant was clearly constitutional.

Oddly, Justice Stevens made his *Garrison* argument appear much more difficult than it actually was. At the outset, he divided the question of the validity of the search and seizure into two separate questions. First, he assessed the validity of the warrant in view of the error about the number of third-floor apartments. Second, he addressed the constitutionality of the execution of the warrant in view of the police error.[339]

Justice Stevens had no difficulty answering the first question.[340] He concluded the police error did not invalidate the warrant because "[t]he validity of the warrant must be assessed on the basis of the information that the officers disclosed, or had a duty to discover and disclose, to the issuing Magistrate."[341] He states "There is no question that the warrant was valid and was supported by probable cause."[342] This conclusion is sound[343] (assuming one does not share Justice Blackmun's view of probable cause).

---

339. Justice Stevens clearly disconnects the two issues. He states, "In our view, the case presents two separate constitutional issues, one concerning the validity of the warrant and the other concerning the reasonableness of the manner in which it was executed." 480 U.S. at 84 (citing Dalia v. United States, 441 U.S. 238, 258 (1979)). *Dalia*, however, does not require this bifurcation of the inquiry; it merely indicates that the execution of a warrant may give rise to a question regarding the reasonableness of the police conduct.

340. He posed the first inquiry as "whether that factual mistake [about there being more than one apartment on the third floor] invalidated a warrant that undoubtedly would have been valid if it had reflected a completely accurate understanding of the building's floor plan." 480 U.S. at 85.

341. *Id.* The reference to the officer's duty in this statement is obscure, however, if it is held up against the Court's earlier holding in Franks v. Delaware, 438 U.S. 154 (1978). *Franks* adopted the rule that a warrant cannot be attacked on the ground that there was false information in the warrant affidavit unless it is shown that officers acted recklessly or with deliberate malice in providing the false information. *Id.* at 155-56. *Franks* does not require the police to behave reasonably (nonnegligently); it only requires that they not be more than negligently careless. *Id.* at 171.

342. 480 U.S. at 81. Oddly, Justice Stevens cast some doubt over this conclusion by writing "the warrant, insofar as it authorized a search that turned out to be ambiguous in scope, was valid when it issued." *Id.* at 85-86. This "when it issued" qualification, however, does not appear to have been meant to imply the warrant later became invalid. Only a few lines earlier Justice Stevens wrote "the discovery of facts demonstrating that a valid warrant was unnecessarily broad does not retroactively invalidate the warrant." *Id.* at 85.

It does not appear Justice Stevens meant the warrant was merely presumptively valid until it was later ruled invalid. His treatment of the validity of the warrant in *Garrison* should not be confused with the Court's treatment of the presumptively valid (but ultimately invalid) search warrants at issue in United States v. Leon, 468 U.S. 897 (1984) and its companion case, Massachusetts v. Sheppard, 468 U.S. 981 (1984).

343. This aspect of *Garrison* is comparable to Justice Scalia's example in *Rodriguez* of a warrant based on a factual inaccuracy. *See supra* note 290.

Justice Stevens then proceeded to the second issue: "whether the execution of the warrant violated [Garrison's] constitutional right to be secure in his home."[344] Oddly, he found the answer to this question "is somewhat less clear" than the question of whether the warrant was valid.[345] It is less clear, however, only because Justice Stevens's opinion lost track of the correct question. He began his discussion of the execution of the warrant in an almost straightforward fashion by writing there is "no difficulty concluding that the officers' entry into the third-floor common area was legal; they carried a warrant for those premises."[346] This is patently correct: if police act pursuant to the terms of a validly issued warrant, their conduct clearly comports with Fourth Amendment reasonableness.[347]

The mystery is why Justice Stevens limited his statement to "the third-floor common area."[348] He had already concluded not only that the warrant's language authorized a search of the entire third floor, but also that there was probable cause for that search. Therefore the police actually carried a valid warrant *for the entire third floor*, including the space that turned out to be Garrison's apartment.[349] The validity of the search in execution of the terms of the valid warrant should not have been less clear; the execution of the warrant was patently constitutional.

Justice Stevens, however, apparently lost sight of that basic point because he treated the police entry of Garrison's apartment under the

---

344. *Garrison*, 480 U.S. at 86.

345. 480 U.S. at 86.

346. *Id.* This statement is followed by Justice Stevens's additional comment that the officers were accompanied by McWebb, who provided the key that they used to open the door giving access to the third-floor common area. *Id.* It does not appear that this fact holds any legal significance for the validity of the entry one way or another; there is no indication that McWebb consented to the entry.

347. The issue that arises most often in a search made in execution of a valid warrant is whether the search exceeded the terms of the search stated in the warrant, but this issue does not arise under the majority opinion's view of *Garrison* because the majority read the warrant as authorizing a search of the entire third floor.

348. 480 U.S. at 86. The validity of the entry into the common area was of particular significance because the State of Maryland claimed the officers could see the drugs in Garrison's apartment in plain view from the common area. This observation, however, would not in itself justify a warrantless entry of Garrison's apartment to seize the drugs. *See* Payton v. New York, 445 U.S. 573, 586-90 (1980). Thus, the validity of the police entry of the common area does not resolve the case. It is not clear, moreover, that the police needed a warrant to be in the common area; there is no indication whether Garrison or McWebb had a reasonable expectation of privacy in the common area. *See* Katz v. United States, 389 U.S. 347 (1967).

349. Justice Scalia describes *Garrison* as a search based on "a warrant supported by probable cause with respect to one apartment which was erroneously issued for an entire floor that was divided (though not clearly) into two apartments." *Rodriguez*, 110 S. Ct. at 2799. This description is not entirely accurate. Justice Stevens clearly concluded the magistrate was presented with probable cause for a search of the entire third floor, not just one apartment.

warrant as though it amounted to a police error in its own right. He then excused that error because it was a "reasonable" factual mistake on the part of the police in light of the "reasonable investigation" they had made.[350] There are two factors that might explain this treatment. One is Justice Stevens's apparent concern with the effect of the police discovery of their error.

Justice Stevens concludes the officers' discovery that there were two third-floor apartments ended any authority for the police to be in Garrison's apartment.[351] This is clearly correct. However, the need for the police to cease searching Garrison's apartment once they learned it was not part of McWebb's apartment does not carry any implication that the police conduct was improper up to that time.[352] This is precisely what *Hill* teaches. The search of Hill's apartment did not become invalid because it was later learned that the wrong person had been arrested. The premise for Justice Stevens's analysis is that there had been no change in the information the police had

---

350. *Garrision*, 480 U.S. at 88.

351. Justice Stevens wrote:

> [T]he validity of the search of [Garrison's] apartment pursuant to a warrant authorizing the search of the entire third floor depends on whether the officers' failure to realize the overbreadth of the warrant was objectively understandable and reasonable. Here it unquestionably was. The objective facts available to the officers at the time suggested no distinction between McWebb's apartment and the third-floor premises.

Id. at 88. He then wrote that the officers were clearly required to limit their search of Garrison's apartment "as soon as they . . . were put on notice of the risk that they might be in a unit erroneously included within the terms of the warrant." *Id.* at 87. He also states "if the officers had known, or should have known," there were two third-floor apartments "they would have been obligated to limit their search to McWebb's apartment." *Id.* at 86-87.

352. The police clearly could not continue to rely on the authority of a warrant once they learned that the description of the place to be searched was overly broad. As soon as they realized this, they were on notice that a continuation of the search would be unreasonable, not merely in the colloquial sense, but because they would have been aware there was no probable cause for the search.

The only claim that might cast any doubt on this analysis is Chief Justice Burger's claim that "Once the warrant issues, there is literally nothing more the policeman can do in seeking to comply with the law." Stone v. Powell, 428 U.S. 465, 498 (1976). That statement was subsequently cited in the rationale for the *Leon* exception to the exclusionary rule. 468 U.S. 897, 921 (1984). One must assume, however, that this statement was invented only to fill a rhetorical gap in the argument for a good-faith exception to the exclusionary rule. It certainly is not an accurate description of the state of the law or of the realities of the warrant process. Despite the mandatory language on search warrants requiring execution and return, there typically is no enforcement of the return requirement and there may be no sanction for an officer's failure to return a warrant. One empirical study reports a substantial proportion of issued warrants are never returned to the issuing court. VAN DUIZAND, *supra* note 17, at 36. A number of courts have explicitly recognized that the police always retain discretion to refrain from searching even after a warrant is issued. *E.g.*, United States v. Alvarez, 812 F.2d 668 (11th Cir. 1987).

regarding probable cause for the warrant as of the time they entered Garrison's apartment. Given that premise, there was no need to treat the execution of the warrant as a separate issue *until* the police discovered their mistake, but that occurred only after they had discovered the drugs. Therefore, the validity of the seizure of the drugs was not compromised by the police discovery of their error.

The other possible reason why Justice Stevens discussed separately the execution of the warrant is that he may have been reluctant to decide the constitutionality of the seizure based only on his interpretation of the scope of the search authorized by the warrant's language.[353] This is suggested by the fact that his discussion of the "reasonableness" of the police mistake,[354] and his citation of the same language from *Brinegar* and *Hill* that Justice Scalia would later cite in *Rodriguez*,[355] occurs in a hypothetical argument that the seizure of drugs from Garrison's apartment would have been valid *even if* the warrant's language authorized only a search of McWebb's apartment.[356] Using two arguments where one would do may seem like a simple example of judicial overkill, but this is an especially inappropriate instance of that vice because the extra argument is not only misdirected,[357] but also fundamentally wrongheaded. A warrant search should not be upheld by claiming the meaning of the language in the warrant stating "the place to be searched" does not really matter. This approach robs the warrant of much of its value as a protection of citizens' privacy.

---

353. This treatment may reflect Justice Stevens's impatience with the formalities of search warrants. *See infra* note 362.

354. Justice Stevens's rhetorical strategy in the majority opinion in *Garrison* is essentially the same as Justice Scalia's in *Rodriguez*. Justice Stevens presents his conclusion as though it were simply an application of a general rule, represented by *Hill*, allowing the police to make "reasonable" mistakes in conducting searches.

355. *Compare* Maryland v. Garrison, 480 U.S. 79, 87 n.11 *with* Illinois v. Rodriguez, 110 S. Ct. 2793, 2800.

Justice Stevens also asserts that "While *Hill* involved an arrest without a warrant, its underlying rationale that an officer's reasonable misidentification of a person does not invalidate a valid arrest is equally applicable to an officer's reasonable failure to appreciate that a valid warrant describes too broadly the premises to be searched." 480 U.S. at 87-88.

356. 480 U.S. at 88. Justice Stevens said "the officers properly responded to the command contained in a valid warrant even if the warrant is interpreted as authorizing a search limited to McWebb's apartment rather than the entire third floor. Under either interpretation of the warrant, the officers' conduct was consistent with a reasonable effort to ascertain and identify the place intended to be searched . . . ." *Id.* Note that Justice Stevens's either-interpretation approach still derives from the language of the actual warrant.

357. Justice Stevens's extra argument does not engage Justice Blackmun's point that Garrison's privacy could not be legitimately breached if the police did not have any cause to think Garrison was involved in selling drugs. *See supra* note 336. From Justice Blackmun's perspective, the search of Garrison's apartment could not have been constitutional regardless of the reason that the police error occurred.

The most significant feature of Justice Stevens's language about "reasonable" police mistakes in *Garrison* is that it sweeps far beyond the kind of error the police actually made. The police error in *Garrison* was reasonable precisely because it was rooted in an understandable inaccuracy in the information that the police proffered to show *probable cause*, and that inaccuracy, in turn, was also incorporated into the language in the warrant that described the place to be searched. So the police error in *Garrison* ran to probable cause just as the error in *Hill* did. Thus, there is no basis to claim the error in *Garrison* evidences a general rule excusing any kind of understandable factual error that the police might make.[358] *Garrison* does not show that a police mistake in executing a warrant should be excused unless it is a mistake that is rooted in the factual information that constitutes the probable cause for the warrant.[359] *Garrison*'s loose statements

---

358. The *Garrision* opinion has not stimulated much commentary, but Professor LaFave's treatment of *Garrision* is instructive. *See* LAFAVE, *supra* note 14, § 4.5, at 30-31 (Supp. 1991). LaFave does not discuss *Garrison* as a broad recognition of the excusability of understandable police errors. Rather, he reads it as an incremental modificaiton of the treatment of search warrants that are based on misperceptions of the number of dwelling units at a particular address. LaFave describes United States v. Santore, 290 F.2d 51 (2d Cir. 1960), as the leading case in the area. In *Santore*, the police believed the building they wished to search was a one family house and they obtained a search warrant on that basis. The police learned when they started the search that the building was a multi-unit dwelling. The court concluded the search warrant was nevertheless valid because it described the premises with the "practical accuracy" required. *Id.* at 67. The reference to practical accuracy is a reference to the probable cause standard as it applies to the place to be searched. In *Santore* the police discovered their error at the outset of the search and made reasonable efforts to confine the search to the specific living unit for which there was probable cause to search.

LaFave views *Garrision* as essentially applying the *Santore* approach although the *Garrison* opinion does not mention *Santore*. LAFAVE, *supra* note 14, at 30 (Supp. 1991). He notes, however, that the situation in *Garrison* is more troubling because the police error occurred in a setting in which the police were aware they were dealing with a multi-unit building. *Id.* Professor LaFave's reading of *Garrison* as an incremental decision is preferable to the sweeping interpretation of *Garrison* evident in *Rodriguez*.

359. *Garrison* is a case in which the police error was rooted in the probable cause assessment and carried over into the language in the warrant. The situation would be quite different in a search in which officers made a mistake and went to premises that were outside both the probable cause showing and the language in the warrant. The following examples illustrate my point.

First, assume an informant is confused about the number of the apartment in which he bought drugs. The number of the apartment he was in was actually "9" but he thought it was "6" because a nail was missing and the number was hanging upside down. The police obtain a warrant, using the informant's information for probable cause, for apartment "6." The police enter the real apartment "6," and completely by chance, find and seize drugs. This seizure would appear to be constitutional because it was conducted pursuant to the express terms of a valid search warrant, notwithstanding that the probable cause showing for the warrant

about "reasonable" police errors are only dicta[360] because there was never any reason for the police execution of the warrant to be treated as a separate issue.[361]

That said, *Garrison* can be viewed as a precursor of *Rodriguez* in the sense that *Garrison* did reveal the majority Justices' disinclination to closely analyze the implications of understandable police errors. The addition of the unnecessary second argument in *Garrison* may even indicate the majority Justices were aggressively seeking

---

incorporated a mistake about the location. This warrant is valid under Franks v. Delaware, 438 U.S. 154 (1978), because there was no "willful or deliberate" misstatement of information by the police. This is the type of error that Justice Scalia gave as an example of a "reasonable" factual error incorporated in a valid warrant. *See supra* note 290.

Second, assume the police obtain a warrant for apartment "6" on the basis of the informant's information, as set out above, but enter apartment "9" where the nail is still missing and the number still looks like "6." This could also be treated as a valid warrant search (assuming it was understandable that the discrepancy between the warrant description and the correct number did not come to the officers' attention before they entered) because the police entered the apartment to which the magistrate was refering when she issued the warrant to search apartment "6." (It is, after all, the apartment to which the informant was referring.) The police error in this hypothetical situation would appear to be comparable to the police error in the entry of Garrison's apartment. The warrant was framed and executed in terms of the flawed information that constituted probable cause. Compare this to Justice Stevens's conclusion that the search warrant in Massachusetts v. Sheppard was valid because there was substantial compliance with the probable cause requirement. 468 U.S. 981, 983-86 (1984) (Stevens, J., concurring).

There is a third type of error, however, which should not be excused. Assume the officers obtain a warrant for apartment "6," as in the previous scenarios. However, they enter apartment "16," because the "1" had fallen off. Apartment "16" is not the apartment identified in the warrant's language. Neither is it the apartment for which the police officers showed probable cause. In a case such as this, which really does involve an error that runs solely to a misexecution of a warrant and does not arise out of the information used for probable cause for the warrant in any way, there is no basis for describing the police mistake as being reasonable within the traditional meaning of the Fourth Amendment (regardless of whether it was understandable for the police to make the error in the circumstances). There is neither valid authority nor probable cause for this last search. Thus, the language in Justice Stevens's *Garrison* opinion should not be read to address the factual scenario set out in this third hypothetical situation because the police error in this scenario is qualitatively different from the police error in *Garrison*.

360. *See* Justice Stevens's conclusion in Horton v. California, 110 S. Ct. 2301, 2307-08 (1990), that statements in Coolidge v. New Hampshire, 403 U.S. 443 (1971), regarding the inadvertence aspect of in plain view observations are not binding precedent because the discussion was not necessary to the decision.

361. Although the execution of the warrant can be thought of as a second stage of police conduct (as distinct from the first stage in which the warrant is applied for and obtained), this does not mean that an error in the information provided to a magistrate necessarily becomes a second and distinct error when the warrant is executed. The legal authority represented by the issued warrant remains in effect unless and until the police learn the information on which the warrant was obtained is incorrect.

opportunities to announce a general amnesty for all manner of police errors about facts (though it is surprising Justice Stevens would take the lead in that project[362]). This proclivity toward amnesty for police errors, however, does not make *Garrison* a legitimate precedent for *Rodriguez*. The language in *Garrison* suggesting broad approval of "reasonable" police errors is as devoid of an analytic rationale or appropriate authority as is the discussion in *Rodriguez*.

C. *The Implications of the "Reasonable" Police Errors About Legal Authority in* Leon *and* Krull

The radical nature of *Rodriguez*'s treatment of police errors about legal authority can be brought into sharper focus by comparing it to the Court's treatment of similar police errors in *Leon* and *Krull*. In both *Leon* and *Krull*, the Court created exceptions to the exclusionary rule. Both cases involved searches conducted pursuant to constitutionally defective legal authority. *Leon* involved a search made pursuant to a warrant that was constitutionally defective because it lacked an adequate showing of probable cause.[363] *Krull* involved a search made pursuant to a statute that was invalid because it did not require any showing of probable cause.[364] The Court concluded the police

---

362. Perhaps the reason Justice Stevens joined the majority in *Garrison* is that he attaches great importance to the fact that the police bothered to obtain a warrant. It is significant that Justice Stevens was the only member of the Court who concluded the warrant search was constitutional in Massachusetts v. Sheppard, 468 U.S. 981 (1984). In *Sheppard*, the warrant affidavit prepared by the police correctly identified the items for which the police were looking, but the obsolete form warrant signed by the magistrate incorrectly listed other items. *Id.* at 986. Justice White's majority opinion assumed the warrant was invalid, but concluded the evidence could be admitted under the *Leon* exception to the exclusionary rule. *Id.* at 990-91. Justice Brennan's dissenting opinion concluded the warrant was constitutionally defective and the evidence therefore had to be suppressed. *Id.* at 986-88. Justice Stevens, however, concluded the warrant search substantially complied with the probable cause and particularity requirements and met constitutional muster. *Id.* at 983-85 (Stevens, J., concurring).

Unfortunately, the language in *Garrison* regarding reasonable police errors is not confined explicitly to the warrant setting.

363. 468 U.S. at 905. The Court did not actually decide the warrant was invalid; it simply assumed the *Leon* warrant was invalid because the Government did not contest that issue. A number of commentators have suggested the warrant in *Leon* was actually valid under the probable cause standard announced the prior term in *Gates*. *E.g.*, Wayne R. LaFave, *The Seductive Call of Expediency: United States v. Leon, Its Rationale and Ramifications*, 1984 U. ILL. L. REV. 895, 911 (1984).

364. 480 U.S. at 342. The search in *Krull* was based on a chop-shop statute that a lower court held to be unconstitutional the day after the *Krull* search occurred. The lower court held the statute unconstitutional because it authorized searches without requiring a showing of probable cause. *Id.* at 345-46. The Supreme Court simply assumed the statute was unconstitutional and therefore did not directly address

officers were not at fault in failing to detect the unconstitutionality of the purported source of legal authority for the search in either case. In both cases, the Court concluded the officers "objectively reasonably relied" upon a presumptively valid (but actually constitutionally defective) source of legal authority for the search.[365]

The notable feature of these cases for present purposes is that no Justice suggested, in either case, that the "objectively reasonable" quality of the police error regarding the validity of the legal authority for the search could conceivably have satisfied the Fourth Amendment. The Justices were unanimous in treating the searches in both cases as being *unconstitutional* despite the "objectively reasonable" quality of the police misperception about the legal authority for the searches. Thus it is evident *Leon* and *Krull* did not adopt the colloquialized notion of Fourth Amendment "reasonableness" espoused in *Rodriguez* (or in the *Garrison* dicta). *Leon* and *Krull* both demonstrate defective legal authority for a search cannot be cured by the quality of the police conduct. If it could, the "objectively reasonable" police conduct in these cases would have made the searches constitutional.[366] The fact that the Court treated the searches as unconstitutional shows that legal authority for an intrusion cannot be derived from a police misunderstanding, even one that is purportedly understandable in the circumstances.

---

that issue. *See id.* at 359.

Subsequent to *Krull*, however, the Court upheld the constitutionality of a similar chop-shop statute in New York v. Burger, 482 U.S. 691 (1987), on the theory that a closely regulated business may be inspected without a warrant under certain conditions. *Id.* at 702.

365. *See Leon*, 468 U.S. at 922; *Krull*, 480 U.S. at 356-57.

366. Neither the *Leon* nor the *Krull* opinion uses the term "objectively reasonable" to refer to Fourth Amendment reasonableness; otherwise those opinions would have concluded the searches were constitutional. The Court uses the term "reasonable" only in the looser and more generic sense of understandable, sensible, nonculpable conduct, as in the colloquial notion of reasonable behavior. Thus, the *Leon* exception can be described as applying to reasonable, unreasonable warrant searches (that is, to reasonable [in the colloquial sense], unreasonable [in the Fourth Amendment sense] searches).

The Court also employs "reasonableness" in the context of civil rights actions for damages for unconstitutional searches in a fashion that is comparable to that found in *Leon* and *Krull*. *See, e.g.*, Anderson v. Creighton, 483 U.S. 635 (1987).

Although the *Rodriguez* majority uses the same unstructured notion of colloquial "reasonableness" found in *Leon*, *Krull*, and *Anderson*, its treatment is novel in using the colloquial notion of "reasonableness" as though it were equivalent to the concept of reasonableness in the Fourth Amendment.

One student commentator has misinterpreted *Rodriguez* as though it were an extension of the *Leon* "reasonable good faith test." *See* David Clark Esseks, Comment, *Errors in Good Faith: The Leon Exception Six Years Later*, 89 MICH. L. REV. 625, 659 n.193 (1990). This mistake is rooted in the Court's utterly inconsistent usage of "reasonable" in relation to stopped police conduct and police error. It is a mistake that is likely to be repeated many times.

Leon and *Krull* confirm the bedrock proposition that legal authority is not probabilistic.[367] They confirm that legal authority for the conduct of government agents is categorical; it either exists or it does not. This is hardly a surprise; the nonprobabilistic quality of legal authority appeared to have been a given prior to *Rodriguez*'s invention of "seeming" authority. Indeed, the nonprobabilistic nature of legal authority had appeared to be the cornerstone of our legal system at least as far back as *Marbury v. Madison*.[368]

Leon and *Krull* also demonstrate the disproportionality of the importance the *Rodriguez* majority attaches to the purported understandableness of a police error concerning legal authority for a search. *Leon* and *Krull* teach that even understandable and blameless police reliance on a *presumptively valid* source of legal authority for a search (a warrant issued by a magistrate or a statute passed by a legislature) cannot suffice to make a search constitutional if the source of legal authority is ultimately found to be defective. How, then, is it conceivable a search could meet constitutional muster if it is predicated solely on an erroneous police assessment that a lay third party (Gail Fischer) possessed legal authority to consent?[369] The Court was right to treat the searches in *Leon* and *Krull* as unconstitutional because there was no valid source of authority. In light of the Court's treatment of the searches in those cases, however, it is incongruous to suggest the mere "seeming" authority of a third party could suffice to make the search in *Rodriguez* constitutional.[370]

One might object, of course, that *Leon* and *Krull* do not deal with factual errors on the part of the police, but this distinction is hardly clear cut. As noted above, Justice Scalia's labeling of the police error in *Rodriguez* as factual is arbitrary. It is not difficult,

---

367. Legal decisions speak in terms of probable cause, but not in terms of probable authority. In other words, we readily distinguish between a mere appearance of authority and authority. The very fact that we draw that distinction indicates authority (and consent) are not understood to be probabilistic in nature. The concept of apparent authority in the law of agency is instructive: it is not merely an appearance of authority but rather constitutes a form of binding authority. *See supra* note 87.

368. Marbury v. Madison, 5 U.S. (1 Cranch) 137 (1803).

369. The State of Illinois attempted to shoehorn the alleged police error in *Rodriguez* into the language of the *Leon* exception by arguing that the officers "acted in reasonable, good faith reliance on Gail Fischer's apparent authority to permit their entry." *See supra* note 65. Although the police are entitled to treat warrants issued by a magistrate or a judge as presumptively valid, the police officers were not entitled to treat a girlfriend's (Fischer's) consent as presumptively valid. It is only the judicial origin of a warrant that entitles it to a presumption of validity.

370. The facts in *Rodriguez* are not susceptible to the *Leon* or *Krull* exceptions to the exculsionary rule. *See supra* note 65. It is entirely possible, however, that the majority would have recognized a broad mistake exception to the exclusionary rule if they had not chosen to recognize "seeming consent" as a basis for a constitutional search.

moreover, to imagine other factual errors that would run to legal authority. Consider this hypothetical scenario: A prankster pretends to be an out-of-town magistrate sitting in for a magistrate who is on vacation. The prankster puts on such a convincing performance that police officers, believing the prankster to be a visiting magistrate, submit a search warrant application for a search of a suspect's home to the prankster. The prankster signs the warrant, and the police conduct a search of the home pursuant to the terms of the warrant.

Is it conceivable the Court would say this hypothetical search satisfied the Fourth Amendment because it was "reasonable" the officers were misled in light of the convincing nature of the prankster's performance? It is hard to believe this thought could be seriously entertained.[371] Yet the police error in this scenario would fall under the "general rule" alluded to in *Rodriguez*. The error is quite analogous to the error which Justice Scalia assumes the police made in *Rodriguez*. In both instances, the officers made a factual mistake about a person's status and, on that basis, reached an incorrect conclusion about the person's legal authority. The mere labeling of a police mistake as "factual" is not a sufficient basis for treating an intrusion as constitutional.

D. The Kinds of Police Error That Should Be Excusable Under Fourth Amendment Reasonableness

What, then, should be the appropriate scope of forgiveness for understandable police errors that affect the justification for a police intrusion? The classification of police errors as factual versus legal is inadequate. It is true that any mistake that cannot be treated as factual by any stretch of the imagination, such as police reliance on a constitutionally defective statute, is clearly beyond constitutional salvation. The characterization of an error as "factual," however, is not sufficient to make an error excusable. The better criterion is to judge police errors according to the nature of the determination to which they are relevant. The excusability of police errors under the Fourth Amendment should be determined according to the component of Fourth Amendment reasonableness to which the error is relevant. Police mistakes about factual matters that pertain to the assessment of probable cause or exigent circumstances should be excused if they are understandable under the circumstances, assuming that a substantial notion of understandableness is employed.[372] This treatment is

---

371. The Court has previously indicated that the constitutionality of a warrant depends on whether the person who issued the warrant was legally authorized to do so. *See* Shadwick v. City of Tampa, 407 U.S. 345 (1972).

372. Police assessments of facts pertaining to exigency should also be excusable because exigency is necessarily a probabilistic standard similar to probable cause,

acceptable because the probable cause legal standard is satisfied by a probablistic assessment (as *Brinegar*, *Hill*, and the first argument in *Garrison* demonstrate). On the other hand, police mistakes that run to the legal authority for a search should not be excused even if they can be labeled factual. Errors running to legal authority should not be excused because legal authority calls for a categorical determination, not a probablilistic assessment (as *Leon* and *Krull* demonstrate).[373]

Unfortunately, the expansive "general rule" referred to in *Rodriguez* appears to excuse police errors that run to aspects of legal authority for police intrusions as though there were a "probable authority" standard. The only limit on the reach of *Rodriguez*'s tolerance of police errors is the creativity that prosecutors and lower court judges show in labeling police errors "factual." *Rodriguez* may well turn out to be a major step toward a generalized doctrine of forgiveness of all kinds of "understandable" police errors under the colloquialized notion of "reasonableness."[374]

In summary, the general rule that Justice Scalia asserts is rooted in an arbitrary characterization of the nature of the police error and

---

because it creates a situation in which the officer must make a decision on the basis of her understanding of the factual situation at the time. A police assessment of exigent circumstances not only is dependent upon an assessment of probable cause for the intrusion, but also is itself a form of probable cause assessment. *See, e.g.*, Minnesota v. Olson, 110 S. Ct. 1684, 1690 (1990); Warden v. Hayden, 387 U.S. 294 (1967). Thus, it is appropriate to excuse an understandable police factual error running to an assessment of exigency the same way in which an understandable police error about probable cause is excused.

373. The criterion I suggest indicates police errors generally should not be excused if they run to issues regarding whether there was consent for an intrusion, whether an intrusion implicated a citizen's reasonable expectation of privacy (that is, whether the Fourth Amendment applies to the intrusion), or whether an intrusion falls within one of the recognized exceptions to the warrant requirement. Errors running to these determinations should not be excusable because each of these determinations essentially involves the existence of legal authority for an intrusion, not the assessment of cause for one.

The significant exception to this proposition is that understandable police errors about facts should be excused if they run to the showing of an exigency that justifies the nonapplication of the warrant requirement. The reason for this exception is the special character of a determination of exigency; an exigent circumstance is treated as an exception to the warrant requirement because it is by definition a circumstance in which it is necessary for the police to act promptly and in which it is infeasible for them to secure a warrant without seriously jeopardizing public interests. The necessity for police action confers temporary legal authority on police officers to decide whether to intrude. *See supra* note 372.

374. Although *Rodriguez* rests on a different doctrinal footing than the proposal for a broad good-faith mistake exception to the exclusionary rule reflected, in part, in the exceptions created in *Leon* and *Krull*, the practical implication is essentially the same. As a practical matter the ultimate issue litigated in search cases is whether seized evidence is admissible, and *Rodriguez*, *Leon*, and *Krull* each increase the likelihood that it will be.

in a misdescription of precedent. Properly understood, *Brinegar*, *Hill*, and *Garrison* stand only for the precise (and noncontroversial) proposition that police mistakes of fact that are understandable under the circumstances do not violate Fourth Amendment reasonableness if they run to the showing of probable cause. None of these cases suggests any basis, however, for excusing police errors that relate to the legal authority for an intrusion. Moreover, *Leon* and *Krull* teach that police errors regarding legal authority violate the Fourth Amendment regardless of the understandable quality of the police conduct.

The *Rodriguez* majority's overly forgiving treatment of police errors is an outgrowth of its refusal to acknowledge the traditional concept of Fourth Amendment reasonableness and the different functions that the probable cause requirement and the legal authority requirement play in that conceptual scheme. The majority's willingness in *Rodriguez* to excuse a police error running to legal authority demonstrates the slippery slope inherent in the Court's embrace of colloquialized "reasonableness."

VII. CONCLUSION

I have engaged in this detailed criticism of the majority's rationale for *Rodriguez* because I think it reveals the degree to which the majority Justices are prepared to disregard or distort doctrine to uphold dubious government intrusions of citizens' privacy. *Rodriguez* trivializes the Fourth Amendment. From beginning to end it reflects an ideological hostility to the values evident in the traditional understanding of the Fourth Amendment. Justice Marshall was clearly correct when he observed *Rodriguez* "tak[es] away some of the liberty the Fourth Amendment was designed to protect."[375]

In this concluding section, I discuss some of the implications of the Rehnquist Court's cavalier treatment of "reasonableness." I first speculate about what *Rodriguez* indicates about the Court's attitude toward consent intrusions, colloquialized "reasonableness," and police mistakes. I then offer a few thoughts about the larger question of what *Rodriguez* portends for the fate of Fourth Amendment doctrine at the hands of the Rehnquist Court.

One implication that clearly emerges from the contorted rationale offered in *Rodriguez* is that the Rehnquist Court will go to great lengths to make it as easy as possible for the police to claim there was "consent" for an intrusion. The Court's eagerness to facilitate findings of consent is not new. The same attitude is evident in the Burger Court's refusal in *Schneckloth* to require officers to inform consenting parties of their right to withhold consent, although

375. 110 S. Ct. at 2807 (Marshall, J., dissenting).

Schneckloth still required that consent be given in fact.[376] The *Rodriguez* majority's nonchalance about who gives "consent" evidences even less regard for citizens' privacy than did *Schneckloth*. If one reads *Rodriguez* in conjunction with the 1991 decisions in *Bostick* (which pushes the notion of "voluntary" consent into the realm of fantasy[377]) and *Jimeno* (which adopts an expansive approach to interpreting the scope of consent[378]), it is fair to say it is unlikely the Rehnquist Court will meet many consent claims that it does not like.

There is a legitimate place for consent in Fourth Amendment analysis. Consent standards that are too loose, however, amount to an end-run around the core meaning of the Fourth Amendment—the prohibition against the government intruding on citizens' privacy unless it has cause to do so. A finding of consent—and now even of "seeming consent"—negates that prohibition. The detriment of too easy consent claims is especially pronounced when they are used to justify highly invasive intrusions of citizens' homes. The Rehnquist Court's easy attitude toward consented searches and even "seemingly consented searches" places a large proportion of police intrusions beyond even the cursory review of cause provided under the Court's colloquialized version of Fourth Amendment reasonableness.

The second implication evident in *Rodriguez* is that the Court has embraced the previously rejected generalized reasonableness reading of the Fourth Amendment. The shortcoming of the loose, colloquialized notion of "reasonableness" applied in *Rodriguez* is apparent; it means almost anything one wants it to. It amounts, in practice, to a rhetorical cover under which statist judges will find it easy to justify ever-expanding police powers.[379] Under colloquialized "reasonableness" the possibilities for justifying searches are greatly expanded. Unfortunately, this likely is precisely the characteristic that makes colloquialized reasonableness so highly prized in the Rehnquist Court.

It is wishful thinking to hope the *Rodriguez* majority's trivialization of Fourth Amendment reasonableness is merely a one shot departure to overcome the ideosyncratic doctrinal obstacle posed by the traditional concept of consent. It is far more likely that *Rodriguez*'s loose treatment of "reasonableness" will be exploited in future decisions that will further undermine the traditional understanding of the meaning of the Fourth Amendment in the context of highly invasive police intrusions.[380] It will not be surprising if future decisions

---

376. *See supra* note 112, 119, and accompanying text.

377. *See supra* text accompanying notes 258-64.

378. *See supra* note 149.

379. Abstractly, it might seem that a contentless standard could be applied in either direction. Practically, however, institutional tendencies are such that discretionary standards favor the state. *See supra* note 9.

380. In fact, Chief Justice Rehnquist already has provided a preview of the

cite *Rodriguez* as authority for the proposition that probable cause is not always a requisite for a police intrusion of a home.

The third tendency evident in *Rodriguez* flows directly from the colloquialized treatment of "reasonableness." The Rehnquist Court majority can be counted on to give police officers every conceivable benefit of the doubt by readily excusing otherwise unconstitutional intrusions as understandable police mistakes.[381] The *Rodriguez* majority's willingness to give police officers a wide berth for mistakes is evident in it sweeping toleration of factual inaccuracies and in its willingness to characterize a mistake about legal authority as a factual determination.

The current Court will not condemn searches and exclude evidence unless "the constable has *flouted* the Fourth Amendment."[382] *Rodriguez*'s ready toleration of police error reflects the triumph of the "regulatory canon"[383] over the traditional understanding that the Fourth Amendment conveys enforceable rights to citizens.

The final tendency of the Rehnquist Court evident in *Rodriguez* is the most disturbing. The Court appears to be determined not to let doctrine interfere with expansive police powers. In other words, the Court seems to be determined not to take Fourth Amendment doctrine seriously. This posture has profound implications.

Constitutional rights are inherently fragile and vulnerable in large part because the enforcement of such rights is entrusted to courts that are, in every real sense, extensions of the very government whose power is meant to be bounded. Given this institutional reality, rights can have meaning only to the degree they are supported by coherent doctrine and are given operative meaning in rules. Legal formalism is a pragmatic requisite for the meaningful articulation of rights.[384]

The Justices of the Rehnquist Court are not ignorant of the relationship between formal doctrine and the enforceability of rights.[385]

---

conclusory fashion in which the majority may employ *Rodriguez* as authority for generalized reasonableness in the future. He has cited *Rodriguez* as authority for the criterion-begging statement that: "The Fourth Amendment does not proscribe all state-initiated searches and seizures; it merely proscribes those which are unreasonable." Florida v. Jimeno, 111 S. Ct. 1801, 1803 (1991) (citing *Rodriguez*).

381. The Court seems to draw the line for acceptable police conduct at a gross recklessness standard. Merely negligent police misconduct is reclassified as "objectively reasonable," which is a rather unusual usage of that term. For example, each of the four limits on the *Leon* exception is stated in a way that indicates objectively reasonable reliance is violated only by conduct that is worse than merely negligent misconduct. *Leon*, 468 U.S. at 923.

382. Kamisar, *Remembering the Old World*, *supra* note 2, at 555 (emphasis added); *see also* Yackle, *supra* note 2, at 423 (Burger Court limits exclusion to evidence seized during flagrantly abusive searches).

383. *See supra* note 7.

384. *See supra* note 9.

385. *See, e.g.*, Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 U.

To the contrary, the pattern that appears in the Court's recent pronouncements is that the Court is consciously engaged in undoing precisely those aspects of Fourth Amendment doctrine that still present an enforceable right. Thus, *Rodriguez* rejects the basic axiom that consent can only arise from the conduct of a person whose interest is at stake—a root principle of Anglo-American jurisprudence. It also rejects the central idea of the Fourth Amendment: the "reasonableness" of an intrusion must be assessed in terms of cause for the intrusion. Indeed, *Rodriguez* even blurs the categorical nature of legal authority. *Rodriguez* rejects the entire project of defining Fourth Amendment rights through legal doctrine.[386]

The style in which Justice Scalia announces the rationale for *Rodriguez* is nearly as disturbing as the content of the decision. The majority opinion lacks candor. Justice Scalia does not write that traditional doctrine is inadequate or outdated and needs changing. He does not explain the reasons for adopting the positions that the majority adopts. He merely asserts radical claims while he pretends they are nothing new and engages in a series of rhetorical obfuscations and diversions.[387] There is no judicial restraint, strict construction or original intent in the *Rodriguez* opinion. There is only a sort of "we've got the votes so we can say anything we want" hubris.

In the final opinion he announced before retiring from the Court, Justice Marshall complained that "Power, not reason, is the new currency of this Court's decisionmaking."[388] That description is also apt for the majority opinion in *Rodriguez*. *Rodriguez* may be a binding interpretation of the Fourth Amendment in the sense that lower courts must follow it. It is, however, neither candid nor sound.

CHI. L. REV. 1175, 1187 (1989) (acknowledging totality of the circumstances standards are nonrules that simply confer fact-finding discretion on reviewing courts and calling for rules to be extended as far as the nature of the question allows).

386. This inclination is not unique to the Rehnquist Court, however; it is also evident in Burger Court search decisions. *See* Yackle, *supra* note 2, at 336, 391, 427-28.

387. *Cf.* Alan M. Dershowitz & John Hart Ely, Harris v. New York: *Some Anxious Observations on the Candor and Logic of the Emerging Nixon Majority*, 80 YALE L.J. 1198 (1971). The authors assert that an opinion that changes prior doctrine should "acknowledge that it was working important changes in the law" and give the issue "substantial treatment"; it should not simply announce "a vote—a reflection of numerical power." *Id.* at 1226.

388. Payne v. Tennessee, 111 S. Ct. 2597, 2619 (1991) (Marshall, J., dissenting).

ARTICLE II OF THE UNIFORM PROBATE CODE AND THE MALPRACTICE REVOLUTION

MARTIN D. BEGLEITER*

Perhaps the two leading developments in the law of wills and trusts over the last thirty years have been the vast increase in the number of lawsuits alleging legal malpractice in an estate planning context[1] and the development of the Uniform Probate Code (UPC).[2] Since 1961, the number of malpractice actions involving aspects of estate planning has skyrocketed.[3] Attorneys have become increasingly concerned with the issue, and articles and continuing legal education programs on the subject abound.[4] The UPC has been adopted in fifteen states and substantially adopted in at least one other state,[5] but the Code's influence extends much further. When a state revises its probate laws, consideration of the UPC provisions is almost automatic, and even if a state does not adopt the entire UPC, the

---

* Richard M. and Anita Calkins Distinguished Professor of Law, Drake University Law School. B.A., 1967, University of Rochester; J.D., 1970, Cornell University. The author expresses his appreciation to Professor Amy Morris Hess of the University of Tennessee College of Law for her helpful comments on an earlier draft of this article. The author wishes to express his gratitude to Laurie A. Butler and Jeffrey Ramsey, Drake University Law School Class of 1993, for their valuable assistance in the research and preparation of this Article. The author gratefully acknowledges the assistance of the Board of Governors of the Drake University Law School Endowment Trust and of Dean David S. Walker of Drake University Law School for the award of a Drake Law School Endowment Trust Research Stipend, which greatly aided the preparation of this article.

1. Martin D. Begleiter, *Attorney Malpractice in Estate Planning—You've Got to Know When to Hold Up, Know When to Fold Up*, 38 KAN. L. REV. 193 (1990) [hereinafter Begleiter].

2. UNIFORM PROBATE CODE (1990). The National Conference of Commissioners on Uniform State Laws and the American Bar Association approved the UPC in August 1969.

3. Begleiter, *supra* note 1, at 263-64 and n.478. *See also* Gerald P. Johnston, *Avoiding Malpractice Claims That Arise Out of Common Estate Planning Situations*, 63 TAXES 780 (1985); *see generally* 2 RONALD E. MALLEN & JEFFREY M. SMITH, LEGAL MALPRACTICE §§ 19.27, 26.1-.10 (3d ed. 1989) [hereinafter MALLEN & SMITH].

4. *See* Begleiter, *supra* note 1, at 273-74 and nn.524-27.

5. UNIFORM PROBATE CODE, 8 U.L.A. 1 (Supp. 1991). The states adopting the UPC are Alaska, Arizona, Colorado, Florida, Hawaii, Idaho, Maine, Michigan, Minnesota, Montana, Nebraska, New Mexico, North Dakota, South Carolina and Utah. Alabama has substantially adopted the UPC. *Id.*

state may enact many provisions of it.[6] Moreover, most law school casebooks devote substantial discussion to the UPC and use its provisions as illustrations of various topics.[7] Students thus become familiar with the UPC and may espouse its theories in the courtroom, the legislature or on state Bar Association committees. The influence of the UPC is both significant and widespread.

In 1990, Article II of the UPC underwent substantial revision as a result of a detailed study undertaken by the Joint Editorial Board for the Uniform Probate Code and a special drafting committee to revise Article II.[8] A number of major changes were made to the provisions of Article II. The revisions (hereinafter referred to as the "new UPC") were approved by the National Conference of Commissioners on Uniform State Laws in July 1990.

This article will examine the impact of some of the major changes made in Article II of the UPC and, in certain cases, minor changes made in particular sections of Article II, to determine whether the provisions are likely to result in an increase or decrease in malpractice litigation. A word of caution is in order. Very little, if any, debate occurred concerning the effect of the changes on malpractice litigation during the revision process. The changes were made because they represented, in the opinion of the Joint Editorial Board and the special Drafting Committee, the most carefully crafted provisions implementing the most widely held public policy.[9] Nevertheless, it is appropriate to examine the revised Article II to evaluate the likelihood of the new provisions leading to malpractice litigation or diminishing the chances that malpractice actions will be instituted. Before under-

---

6. This was apparently the case in the recent revisions of the probate laws of New Jersey and California. *See* Harrison F. Durand, *New Jersey Adopts the Uniform Probate Code,* 9 PROB. & PROP. 10 (no. 2, 1980); 20 CAL. L. REV. COMM. RPT. 1001 (1990) (California).

7. *See, e.g.*, EUGENE F. SCOLES & EDWARD L. HALBACH, JR., PROBLEMS AND MATERIALS ON DECEDENTS' ESTATES AND TRUSTS 47-48 (intestacy), 93-96 (elective share) (4th ed. 1987); LAWRENCE W. WAGGONER, ET AL., FAMILY PROPERTY LAW xix (1991) (casebook "centers on the Uniform Probate Code"); JOHN RITCHIE, ET AL., DECEDENTS' ESTATES AND TRUSTS 89-90 (intestacy), 161-64 (elective share) (7th ed. 1988).

8. UNIFORM PROBATE CODE, article II, prefatory note (1990). The author was the American Bar Association Advisor to the special Drafting Committee to Revise Article II.

9. This is not to say there was agreement over most provisions. There was a great deal of debate over many provisions, and almost every change went through several drafts. The remarkable thing about the process was the general agreement among the Joint Editorial Board, the special Drafting Committee and the Advisors over the aim and general direction of the revisions. The meetings were conducted in a congenial atmosphere, with everyone contributing toward the effort to produce the best possible product. Much of the credit for this remarkable effort must go to Professor Lawrence W. Waggoner, Reporter, and Professor Richard V. Wellman, Educational Director.

taking that examination, a summary of the malpractice revolution is in order.

I. THE MALPRACTICE REVOLUTION: PRIVITY AND THE STATUTE OF LIMITATIONS[10]

A. Privity

Originally, an attorney was liable for malpractice only to his client.[11] Because beneficiaries under a will had no contractual relationship with the attorney who drafted the will, they had no standing to bring an action for malpractice against the drafter.[12] This rule was challenged and abruptly changed in 1961 in the leading case of *Lucas v. Hamm*.[13] The court determined that henceforth, the determination of whether a beneficiary under a will would have standing to sue the drafter would be made as a matter of policy rather than on the presence or absence of privity[14] and discussed the factors involved in the policy determination:

> In restating the rule it was said that the determination whether in a specific case the defendant will be held liable to a third person not in privity is a matter of policy and involves the balancing of various factors, among which are the extent to which the transaction was intended to affect the plaintiff, the foreseeability of harm to him, the degree of certainty that the plaintiff suffered injury, the closeness of the connection between the defendant's conduct and the injury, and the policy of preventing future harm.
>
> . . .
>
> Since defendant was authorized to practice the profession of an attorney, we must consider an additional factor not present in *Biakanja*, namely, whether the recognition of liability to beneficiaries of wills negligently drawn by attorneys would impose an undue burden on the profession.[15]

The court easily disposed of the first four factors in favor of the beneficiaries and ruled no undue burden on attorneys would result

---

10. For an extensive discussion of these problems, see Begleiter, *supra* note 1, at 194-218.

11. 1 MALLEN & SMITH, *supra* note 3, at 360, 366; 2 MALLEN & SMITH, *supra* note 3, at 595; Johnston, *supra* note 3, at 782.

12. Buckley v. Gray, 42 P. 900, 902 (Cal. 1895).

13. 364 P.2d 685 (Cal. 1961), *cert. denied*, 368 U.S. 987 (1962). *Lucas* was preceded by Biakanja v. Irving, 320 P.2d 16 (Cal. 1958), allowing a malpractice suit by a beneficiary against a notary public who prepared an invalid will, despite the lack of privity of contract.

14. *Lucas*, 364 P.2d at 687.

15. *Id.* at 687-88.

from its decision, especially when compared to the loss to the innocent beneficiaries that would result from denying the cause of action.[16] The court further held the beneficiaries had a cause of action as third-party beneficiaries against the attorney for breach of contract.[17] The court justified this decision on policy grounds.[18]

After *Lucas* the beneficiary had two causes of action: one in tort and one in contract. The court in *Heyer v. Flaig* refined this analysis, stating the contract theory was superfluous because there could be no recovery without negligence.[19]

Since 1961, a number of courts have faced the question of whether to retain the privity requirement. Only Nebraska,[20] New York,[21] Texas,[22] Virginia[23] and Ohio[24] continue to require privity. The remaining jurisdictions have dispensed with it.[25] Although one court

---

16. *Id.* at 688.

17. *Id.* at 688-89.

18. *Id.* at 689. The court held, however, that violation of the rule against perpetuities did not constitute negligence by the attorney. *Id.* at 690-91. For a discussion of this aspect of the case, see *infra* Part II C.

19. 449 P.2d 161, 164 (Cal. 1969).

20. Lilyhorn v. Dier, 335 N.W.2d 554 (Neb. 1983); St. Mary's Church of Schuyler v. Tomek, 325 N.W.2d 164 (Neb. 1982).

21. Deeb v. Johnson, 566 N.Y.S.2d 688 (App. Div. 1991), Victor v. Goldman, 344 N.Y.S.2d 672 (Sup. Ct. 1973), *aff'd mem.*, 351 N.Y.S.2d 956 (App. Div. 1974); Maneri v. Amodeo, 238 N.Y.S.2d 302 (Sup. Ct. 1963). Two recent New York cases, however, indicated support for a modification of the privity rule but believed themselves bound by decisions of higher courts. Baer v. Broder, 436 N.Y.S.2d 693 (Sup. Ct. 1981), *aff'd on other grounds*, 447 N.Y.S.2d 538 (App. Div. 1982); Estate of Douglas, 428 N.Y.S.2d 558 (Sur. Ct. 1980); *but see* Viscardi v. Lerner, 510 N.Y.S.2d 183 (App. Div. 1986).

22. Berry v. Dodson, Nunley & Taylor, 717 S.W.2d 716 (Tex. Ct. App. 1986), *judgment set aside by agreement of parties*, 729 S.W.2d 690 (Tex. 1987).

23. Copenhaver v. Rogers, 384 S.E.2d 593 (Va. 1989).

24. Simon v. Zipperstein, 512 N.E.2d 636 (Ohio 1987) (except in cases of fraud, collusion, malice and bad faith).

25. Rathblott v. Levin, 697 F. Supp. 817 (D.N.J. 1988); Wisdom v. Neal, 568 F. Supp. 4 (D.N.M. 1982); Fickett v. Superior Court, 558 P.2d 988 (Ariz. Ct. App. 1976); Krawczyk v. Stingle, 543 A.2d 733 (Conn. 1988); Stowe v. Smith, 441 A.2d 81 (Conn. 1981); Licata v. Spector, 225 A.2d 28 (Conn. C.P. 1966); Needham v. Hamilton, 459 A.2d 1060 (D.C. 1983); McAbee v. Edwards, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976); McLane v. Russell, 512 N.E.2d 366 (Ill. App. Ct. 1987), *aff'd*, 546 N.E.2d 499 (Ill. 1989); Ogle v. Fuiten, 445 N.E.2d 1344 (Ill. App. Ct. 1983), *aff'd*, 466 N.E.2d 224 (Ill. 1984); Walker v. Lawson, 526 N.E.2d 968 (Ind. 1988) (attorney's actions held competent as a matter of law because attempt to accomplish testator's objective would have been contrary to established law, thus attorney had no alternative); Succession of Killingsworth, 292 So. 2d 536 (La. 1974); Flaherty v. Weinberg, 492 A.2d 618 (Md. 1985); Marker v. Greenberg, 313 N.W.2d 4 (Minn. 1981); Albright v. Burns, 503 A.2d 386 (N.J. Super. Ct. App. Div. 1986); Stewart v. Sbarro, 362 A.2d 581 (N.J. Super. Ct. App. Div. 1976); Jenkins v. Wheeler, 316 S.E.2d 354 (N.C. Ct. App. 1984), *reh'g denied*, 3215 S.E.2d 136 (N.C. 1984); Hale v. Groce, 744 P.2d 1289 (Or. 1987); Guy v. Liederbach, 459 A.2d 744

has limited the allowable theory of recovery to breach of contract and the allowable plaintiffs to intended third-party beneficiaries,[26] and three states have incorrectly restricted recovery in malpractice to situations in which the testator's intent is evident on the face of the will or the attorney admits negligence,[27] most courts allow actions in either tort or breach of contract or both.[28]

B. Statute of Limitations

Before 1969 attorneys often defended malpractice actions in estate planning situations by relying on the statute of limitations. More

---

(Pa. 1983); Ward v. Arnold, 328 P.2d 164 (Wash. 1958); Auric v. Continental Casualty Co., 331 N.W.2d 325 (Wis. 1983). *Cf.* Levin v. Berley, 728 F.2d 551 (1st Cir. 1984) (malpractice in estate planning case decided on statute of limitations grounds but implies that privity is not necessary under Massachusetts law); Shideler v. Dwyer, 417 N.E.2d 281 (Ind. 1981) (malpractice in estate planning case decided on statute of limitations grounds but implies that privity is not necessary); Kirgan v. Parks, 478 A.2d 713 (Md. Ct. Spec. App. 1984) (court discussed abandoning the privity requirement but did not decide the question), *cert. denied*, 484 A.2d 274 (Md. 1984); Jaramillo v. Hood, 601 P.2d 66 (N.M. 1979) (malpractice in estate planning case decided on statute of limitations grounds but implies that privity is not necessary); Persche v. Jones, 387 N.W.2d 32 (S.D. 1986) (court adopted reasoning of attorney malpractice cases to hold that beneficiaries could maintain a cause of action against a banker who drafted decedent's will for negligent preparation of a testamentary instrument despite lack of privity).

26. Guy v. Liederbach, 459 A.2d 744 (Pa. 1983). This case is discussed in Begleiter, *supra* note 1, at 204-06.

27. *See* Lorraine v. Grove, Ciment, Weinstein & Stauber, 467 So. 2d 315 (Fla. 1985); DeMaris v. Asti, 426 So. 2d 1153 (Fla. Dist. Ct. App. 1983); Schreiner v. Scoville, 410 N.W.2d 679 (Iowa 1987) (court allowed exception to the rule where events closely connected in time; extent of exception undetermined); Kirgan v. Parks, 478 A.2d 713 (Md. Ct. Spec. App.), *cert. denied*, 484 A.2d 274 (Md. 1984). These decisions are incorrect because the courts fail to realize that in the malpractice action it is the attorney-client contract, not the will, that they are interpreting. The cases are discussed in Begleiter, *supra* note 1, at 198-204, 260-63.

28. Levin v. Berley, 728 F.2d 551 (1st Cir. 1984); Wisdom v. Neal, 568 F. Supp. 4 (D.N.M. 1982); Fickett v. Superior Court, 558 P.2d 988 (Ariz. Ct. App. 1976); Stowe v. Smith, 441 A.2d 81 (Conn. 1981); Licata v. Spector, 225 A.2d 28 (Conn. 1966); Needham v. Hamilton, 459 A.2d 1060 (D.C. 1983) (contract theory not discussed); Lorraine v. Grover, Ciment, Weinstein & Stauber, 467 So. 2d 315 (Fla. Dist. Ct. App. 1985); DeMaris v. Asti, 426 So. 2d 1153 (Fla. Dist. Ct. App. 1983); McAbee v. Edwards, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976); Hamilton v. Powell, Goldstein, Frazer & Murphy, 306 S.E.2d 340 (Ga. Ct. App. 1983), *aff'd*, 311 S.E.2d 818 (Ga. 1984); Ogle v. Fuiten, 445 N.E.2d 1344 (Ill. App. Ct. 1983), *aff'd*, 446 N.E.2d 224 (Ill. 1984); Shideler v. Dwyer, 417 N.E.2d 281 (Ind. 1981); Succession of Killingsworth, 292 So. 2d 536 (La. 1974); Flaherty v. Weinberg, 492 A.2d 618 (Md. 1985); Kirgan v. Parks, 478 A.2d 713 (Md. Ct. Spec. App. 1984), *cert. denied*, 484 A.2d 274 (Md. 1984); Marker v. Greenberg, 313 N.W.2d 4 (Minn. 1981); Stewart v. Sbarro, 362 A.2d 581 (N.J. Super.), *cert. denied*, 371 A.2d 63 (N.J. 1976); Jaramillo v. Hood, 601 P.2d 66 (N.M. 1979); Jenkins v. Wheeler, 316 S.E.2d 354 (N.C. Ct. App.), *reh'g denied*, 321 S.E.2d 136 (N.C. 1984); Hale v. Groce, 744 P.2d 1289 (Or. 1987); Ward v. Arnold, 328 P.2d 164 (Wash. 1958); Auric v. Continental Cas. Co., 331 N.W.2d 325 (Wis. 1983).

specifically, the defense related to the date the cause of action accrued.[29] Under traditional rules a cause of action accrued and the statute of limitations commenced to run when the negligent act occurred or the contract was breached.[30] In wills cases the statute of limitations began to run when the will or other document was executed. Under such a rule the statute of limitations often had expired when the error in the will or document was discovered, which was usually after the testator died.[31] Rarely did the beneficiary have a remedy.

Shortly after dispensing with privity the courts modified the occurrence rule, replacing it with the "injury or damage rule."[32] Under this rule a cause of action does not accrue until injury or damage occurs.[33] This is probably the majority rule in legal malpractice law today.[34] In the case of a will, the application of the rule means that the cause of action does not accrue until the testator's death.[35] Courts reason that because the beneficiaries under a will acquire no legal rights to property before the testator's death, the time at which the testamentary scheme becomes unalterable, they have no injury and no cause of action until then.[36] Only at the testator's death does the cause of action accrue and the statute of limitations begin to run.[37] The court in *Heyer v. Flaig* noted the absurdity of a rule that would permit recovery only for beneficiaries of testators who died within two years of the drafting of the will. Such a rule would be in direct contrast to the policy considerations involved in dismantling the privity barrier, noted in *Lucas v. Hamm*.[38] Other jurisdictions quickly adopted the damage rule.[39]

---

29. A question of which statute of limitations applies to a malpractice action may also be involved. *Compare* Heyer v. Flaig, 449 P.2d 161 (Cal. 1969) (tort) *with* Price v. Holmes, 422 P.2d 976 (Kan. 1967) (tort and contract) *with* Levin v. Berley, 728 F.2d 551 (1st Cir. 1984) (malpractice statute). *See* 2 MALLEN & SMITH, *supra* note 3, at 69-92. Further discussion of the question is beyond the scope of this Article.

30. The leading case stating the rule is Wilcox v. Plummer, 29 U.S. 43, 4 Pet. 172 (1830), which involved an action against an attorney for failure to bring suit to collect on a promissory note. *See* 2 MALLEN & SMITH, *supra* note 3, at § 18.10.

31. Heyer v. Flaig, 449 P.2d 161, 166 (Cal. 1969).

32. *See* 2 MALLEN & SMITH, *supra* note 3, at § 18.11.

33. Heyer v. Flaig, 449 P.2d 161, 165-68 (Cal. 1969); Price v. Holmes, 422 P.2d 976 (Kan. 1967); Succession of Killingsworth, 292 So. 2d 536, 542 (La. 1974). *See* 2 MALLEN & SMITH, *supra* note 3, at §§ 18.11, .18.

34. 2 MALLEN & SMITH, *supra* note 3, § 18.11, at 100.

35. Heyer v. Flaig, 449 P.2d 161, 168 (Cal. 1969).

36. *Id.* at 166.

37. *Id.*

38. 364 P.2d 685 (Cal. 1961).

39. *See, e.g.*, Budd v. Nixon, 491 P.2d 433 (1971); Shideler v. Dwyer, 417 N.E.2d 281 (Ind. 1981); Millwright v. Romer, 322 N.W.2d 30 (Iowa 1982); Price

Some courts have applied other rules in conjunction with the injury or damage rule in malpractice cases. One of these is the "discovery rule," which provides that the cause of action does not accrue until the plaintiff knows or should have known of the injury, its cause, and the defendant's possible negligence.[40] Another is the "continuous representation rule," which defers accrual of the cause of action or tolls the statute of limitations while the attorney continues to represent the client regarding the subject matter in which the error occurred.[41] Development continues to occur as to all these rules as the courts refine statute of limitations doctrine in this area.[42]

II. THE MALPRACTICE REVOLUTION: TYPES OF ERRORS CAUSING MALPRACTICE

Of primary importance in evaluating the effect of the changes in Article II of the UPC on malpractice is a review of the types of errors that have been held to constitute malpractice. A preliminary word of caution is necessary. Many of the cases discussed in this section are appeals from summary judgment motions or from dismissals for failure to state a cause of action. A reversal of such a decision, of course, does not necessarily indicate the court will hold the conduct involved constitutes malpractice. Often a question of fact may result, with the proof adduced on that question determining the outcome. It can at least be said, however, that a failure of the court to rule the conduct was not negligence as a matter of law indicates that if the beneficiary presents a strong enough case, in the right circumstances the court will uphold an award of damages for malpractice.

---

v. Holmes, 422 P.2d 976 (Kan. 1967); Hagen v. Messer, 683 P.2d 1140 (Wash. Ct. App. 1984); Auric v. Continental Cas. Co., 331 N.W.2d 325 (Wis. 1983). *See also* 2 MALLEN & SMITH, *supra* note 3, § 18.11, at 100-02 n.1.

40. Heyer v. Flaig, 449 P.2d 161, 168 n.7 (Cal. 1969); Wall v. Lewis, 393 N.W.2d 758 (N.D. 1986). *See also* Levin v. Berley, 728 F.2d 551 (1st Cir. 1984); Neel v. Maguna, Olney, Levy, Cathcart & Gelfand, 491 P.2d 421 (Cal. 1971); Downing v. Vaine, 228 So. 2d 622 (Fla. Dist. Ct. App. 1969); Millwright v. Romer, 322 N.W.2d 30 (Iowa 1982); Cameron v. Montgomery, 225 N.W.2d 154 (Iowa 1975); Jaramillo v. Hood, 601 P.2d 66 (N.M. 1979); 2 MALLEN & SMITH, *supra* note 3, § 18.14.

41. 2 MALLEN & SMITH, *supra* note 3, § 18.12. *See also* Greene v. Greene, 436 N.E.2d 496, 500-01 (N.Y. 1982), *modified by* McDermott v. Torre, 437 N.E.2d 1108 (N.Y. 1982) (holding the continuous treatment doctrine tolls the statute of limitations rather than delays accrual of the cause of action as held in *Greene*); Wall v. Lewis, 393 N.W.2d 758, 762-65 (N.D. 1986); Brown v. Nichols, No. CA86-10-022, 1987 WL 7594 (Ohio Ct. App. Mar. 9, 1987).

42. For a discussion of the questions remaining in the damage-injury, discovery, and continuous treatment rules, see Begleiter, *supra* note 1, at 210-18.

A. Execution Errors[43]

The majority of estate planning malpractice cases so far decided have involved execution errors. *Auric v. Continental Casualty Company*[44] is typical. Testator executed a will containing a bequest to plaintiff. The drafting attorney signed as one witness. Because of "confusion" or a "mistake of the moment,"[45] however, the attorney's secretary, who was to serve as the second witness, failed to sign the will. The court dismissed the defenses of privity and the statute of limitations, holding the former was not required, and the statute began to run at the testator's death.[46] Because the attorney admitted he negligently supervised the will's execution, the court directed the trial court to enter judgment for the beneficiary.[47] Had the case been appealed from a trial court or jury verdict of negligence, there is little doubt that the verdict would have been upheld. Several other typical mistakes in execution have been held to be malpractice, assuming negligence and causation are proven.[48] In addition, several courts have implied that malpractice liability would result from execution failures in cases ultimately decided on other grounds.[49]

Based on the situations already discussed, a court is highly likely to find any execution error is a basis for imposing malpractice liability. Although not yet decided, cases where two testators (usually husband and wife) execute each other's will[50] and misplaced signature cases (where the testator signs in a place other than at the end of

---

43. For a more extensive discussion of execution errors see Begleiter, *supra* note 1, at 218-22.

44. 331 N.W.2d 325 (Wis. 1983).

45. *Id.* at 327.

46. *Id* at 327, 330.

47. *Id.* at 329.

48. Licata v. Spector, 225 A.2d 28 (Conn. C.P. 1966) (will did not contain signatures of required number of witnesses); Price v. Holmes, 422 P.2d 976 (Kan. 1967) (will not executed in the presence of the attesting witnesses); Persche v. Jones, 387 N.W.2d 32 (S.D. 1986) (testator and witnesses did not sign in the presence of each other). *See also* Guy v. Liederbach, 459 A.2d 744 (Pa. 1983) (beneficiary acted as necessary attesting witness; court adopts third-party beneficiary exception under RESTATEMENT (SECOND) OF CONTRACTS § 302 (1979) to privity and limited malpractice to contract theory).

49. *E.g.*, Bormaster v. Baldridge, 723 S.W.2d 533 (Mo. Ct. App. 1987) (failure to acknowledge trust amendment; summary judgment granted for attorney because statute of limitations had expired); Jaramillo v. Hood, 601 P.2d 66 (N.M. 1979) (execution errors not stated; statute of limitations began running on discovery when beneficiary changed attorneys).

50. *E.g.*, *In re* Estate of Snide, 418 N.E.2d 656 (N.Y. 1981) (court admitted the will actually executed by the deceased husband and reformed it on the ground that what had occurred was so obvious, over a strong dissent emphasizing the possible consequences of the ruling), *In re* Estate of Pavlinko, 148 A.2d 528 (Pa. 1959).

the will in a jurisdiction requiring subscription[51] or where the testator signs a self-proving affidavit but not the will[52]) are prime candidates for malpractice treatment. Similarly, where a court voids an amendment to a trust because the amendment was not acknowledged, the beneficiary should be able to recover in a malpractice action.[53] Indeed, an early case held the attorney subject to liability where the attorney sent the will to the client with adequate instructions for execution, but the client did not follow them correctly.[54] The court believed the failure of the client to correctly follow the instructions should have been anticipated by the attorney and ruled the "sole proximate cause" doctrine inapplicable.[55]

B. Failure to Effectuate Testamentary Desires: Drafting Errors[56]

A number of cases involving relatively simple legal errors have been grouped under the category of errors that frustrated the testator's testamentary desires or drafting errors. Perhaps the clearest instance of this type of error is a case where the attorney erroneously omitted a residuary clause from a will that had been included in several previous drafts of the same person's will.[57] On appeal, the court affirmed summary judgment against the attorney. The court stated expert testimony on the standard of care is not needed when "the attorney's lack of care and skill is so obvious that the trier of fact can find negligence as a matter of common knowledge."[58]

Perhaps the leading case in this area is *Heyer v. Flaig*,[59] which also adopted the rule that the cause of action for malpractice in wills cases accrues on the date of testator's death.[60] Testatrix told defendant attorney that she wished to bequeath her entire estate to her two daughters and that she would soon be married. Before the marriage the attorney drafted a will simply leaving the estate to the daughters, which testatrix executed.[61] The husband claimed a portion of the estate under an omitted spouse provision of California law. The

---

51. *E.g.*, N.Y. EST. POWERS & TRUSTS LAW § 3-2.1 (McKinney 1981). *See In re* Estate of Mergenthaler, 474 N.Y.S.2d 253 (Sur. Ct. 1984).

52. *E.g.*, Boren v. Boren, 402 S.W.2d 728 (Tex. 1966).

53. Bormaster v. Baldridge, 723 S.W.2d 533 (Mo. Ct. App. 1987) (action commenced after expiration of statute of limitations).

54. Ward v. Arnold, 328 P.2d 164 (Wash. 1958).

55. *Id.* at 166.

56. For a more extensive analysis of the type of error see Begleiter, *supra* note 1, at 222-28.

57. Needham v. Hamilton, 459 A.2d 1060 (D.C. 1983); *on appeal from remand*, Hamilton v. Needham, 519 A.2d 172, 174 (D.C. 1986).

58. Hamilton v. Needham, 519 A.2d 172, 174 (D.C. 1986).

59. 449 P.2d 161 (Cal. 1969). A similar case is McAbee v. Edwards, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976).

60. *See* Section I B, *supra*.

61. Heyer v. Flaig, 449 P.2d 161, 162-63 (Cal. 1969).

court held the complaint stated a valid cause of action in malpractice against the attorney, noting a reasonably prudent attorney would "appreciate the consequences of a post-testamentary marriage, advise the testator of such consequences, and use good judgment to avoid them if the testator so desires."[62]

Unless obtaining the testator's desired result is impossible under state law,[63] most errors of this type should provide a solid basis for recovery against the drafting attorney in a malpractice action. Advising a testator that a declaration of the character of property (separate or community) in a will is sufficient to fix the legal ownership of the property without taking other steps has been held to state a cause of action for malpractice.[64] Failure to include a clause requested by testators bequeathing property to plaintiffs if testators died within thirty days of each other, which happened, will support a malpractice action,[65] as will the failure to draft a trust provision or to include a specific bequest in a will in disregard of testator's intent.[66] Although several cases have been cited for the proposition that an attorney is not liable for failing to effectuate testamentary desires,[67] a close reading of these cases shows they either arose in states that retain a strict privity rule,[68] arose in a state that

---

62. *Id.* at 165.

63. *E.g.*, Lorraine v. Grove, Ciment, Weinstein & Stauber, 467 So. 2d 315 (Fla. Dist. Ct. App. 1985) (life estate in homestead could not be bequeathed under Florida law, no allegation or evidence that testator wished to give beneficiary other property if life estate in homestead could not be bequeathed); Walker v. Lawson, 526 N.E.2d 968 (Ind. 1988) (statute and case law did not allow decedent who executed a will shortly before death to prevent husband from electing statutory share or avoid election by lifetime transfer).

64. Garcia v. Borelli, 180 Cal. Rptr. 768 (Ct. App. 1982).

65. Ogle v. Fuiten, 445 N.E.2d 1344, 1346 (Ill. App. Ct. 1983), *aff'd*, 466 N.E.2d 224, 228 (Ill. 1984).

66. Hale v. Groce, 744 P.2d 1289, 1290 (Or. 1987).

67. Ventura County Humane Soc'y for Prevention of Cruelty to Children & Animals v. Holloway, 115 Cal. Rptr. 464 (Ct. App. 1974); Hiemstra v. Houston, 91 Cal. Rptr. 269 (Ct. App. 1970); Lorraine v. Grover, Ciment, Weinstein & Stauber, 467 So. 2d 315 (Fla. Dist. Ct. App. 1985); Kirgan v. Parks, 478 A.2d 713 (Md. Ct. Spec. App. 1984), *cert. denied*, 484 A.2d 274 (Md. 1984); Marker v. Greenberg, 313 N.W.2d 4 (Minn. 1981); Lilyhorn v. Dier, 335 N.W.2d 554 (Neb. 1983); St. Mary's Church v. Tomek, 325 N.W.2d 164 (Neb. 1982); Victor v. Goldman, 344 N.Y.S.2d 672 (Sup. Ct. 1973), *aff'd*, 351 N.Y.S.2d 956 (App. Div. 1974); Maneri v. Amodeo, 238 N.Y.S.2d 302 (Sup. Ct. 1963); Berry v. Dodson, Nunley & Taylor, 717 S.W.2d 716 (Tex Ct. App. 1986), *set aside by agreement of the parties*, 729 S.W.2d 690 (Tex. 1987). *See* Begleiter, *supra* note 1, at 224-28 for a discussion of these cases.

68. Lilyhorn v. Dier, 335 N.W.2d 554 (Neb. 1983); St. Mary's Church v. Tomek, 325 N.W.2d 164 (Neb. 1982); Victor v. Goldman, 344 N.Y.S.2d 672 (Sup. Ct. 1973), *aff'd*, 351 N.Y.S.2d 956 (App. Div. 1974); Maneri v. Amodeo, 238 N.Y.S.2d 302 (Sup. Ct. 1963); Berry v. Dodson, Nunley & Taylor, 717 S.W.2d 716 (Tex. Ct. App. 1986), *set aside by agreement of the parties*, 729 S.W.2d 690 (Tex. 1987).

has adopted the peculiar rule, previously discussed,[69] that the testamentary intention must be apparent on the face of the will to support a malpractice action,[70] or were cases in which the beneficiary failed to allege the true intention of the testator, which is crucial to a malpractice case.[71] None of these cases, properly considered, stand for the proposition that a failure to effectuate the testator's intent is not a valid ground to support a malpractice action.[72]

Based on the results of the cases already decided in this area, it is likely that other common errors, including

> failure to provide protection against pretermitted heirs taking a portion of the estate if a client so desires, . . . faulty creation of a trust, . . . failure to effectuate the testator's wishes regarding children from prior marriages and their issue, . . . [and] creation of irrevocable trusts that were intended to be revocable or vice-versa[73]

are all prime candidates for malpractice liability in the future.[74]

C. *Complicated Legal Errors*

While there have not been many cases involving complicated legal errors, the courts appear willing to sustain malpractice actions in such cases. In *Fazio v. Hayhurst*,[75] a California court upheld a cause of action against an attorney who advised a surviving spouse to elect to take under the will. The attorney prepared a "Decree of Distribution," which incorrectly treated a large amount of community property as separate property, and negligently failed to advise the spouse to revoke her election to take under the will.[76] Although the case was primarily concerned with statute of limitations issues, the opinion suggests that if the allegations contained in the complaint were proved, malpractice liability would attach.[77] Liability also has been imposed against an attorney who drafted a testamentary spendthrift trust with no provision for distribution on the death of the primary beneficiary.[78]

---

69. *See* Section I A, *supra*.

70. Lorraine v. Grover, Ciment, Weinstein & Stauber, 467 So. 2d 315 (Fla. Dist. Ct. App. 1985); Kirgan v. Parks, 478 A.2d 713 (Md. Ct. Spec. App.), *cert. denied*, 484 A.2d 274 (Md. 1984).

71. Ventura County Humane Soc'y for Prevention of Cruelty to Children & Animals v. Holloway, 115 Cal. Rptr. 464 (Ct. App. 1974); Hiemstra v. Houston, 91 Cal. Rptr. 269 (Ct. App. 1970); Marker v. Greenberg, 313 N.W.2d 4 (Minn. 1981).

72. See Begleiter, *supra* note 1, at 225-28.

73. *Id.* at 228.

74. *See id.*

75. 55 Cal. Rptr. 370 (Ct. App. 1966).

76. *Id.* at 370-71.

77. *Id.* at 372.

78. Sizemore v. Swift, 719 P.2d 500 (Or. 1986).

In another leading case[79] liability resulted when an attorney failed to properly prepare a will containing a trust for her son. The corpus of the trust was to be distributed to the beneficiary when he attained age fifty. If the beneficiary died before that age, the corpus was to be paid to his issue.[80] In contrast the trust as drafted by the attorney provided that when the beneficiary attained age fifty, the trust was payable to his issue.[81] The court held the complaint stated a valid cause of action, though noting that proof might be more difficult than in a case where the error appeared on the face of the will.[82]

Surprisingly the most frequently litigated topic in this area is the rule against perpetuities. *Lucas v. Hamm*,[83] the case that overturned privity,[84] held an attorney could not be held liable for malpractice for violating the rule.[85] Stating the rule has "long perplexed the courts and the bar"[86] and referring to the rule as a "technicality-ridden legal nightmare" and a "dangerous instrumentality in the hands of most members of the bar,"[87] the court held an attorney is not negligent as a matter of law for making an error in the perpetuities area.[88] There is reason to doubt that *Lucas* represents the law today. Drafting to avoid the rule no longer seems esoteric;[89] in 1982, the Iowa Supreme Court stated every Iowa resident is presumed to know the rule against perpetuities.[90]

D. Estate Administration[91]

The most frequently litigated problem in this area is malpractice for failure to file tax returns or failure to file returns on time. Almost every court considering such a case has held this type of error is sufficient to support a malpractice action.[92] Most other cases in this

---

79. Stowe v. Smith, 441 A.2d 81 (Conn. 1981).
80. *Id.* at 82.
81. *Id.*
82. *Id.* at 83-84.
83. 364 P.2d 685 (Cal. 1961), *cert. denied,* 368 U.S. 987 (1962).
84. *Lucas,* 364 P.2d at 687-88. *See also supra* Section I A.
85. *Id.* at 690-91.
86. *Id.* at 690, *citing* JOHN CHIPMAN GRAY, THE RULE AGAINST PERPETUITIES xi (4th ed. 1942); W. Barton Leach, *Perpetuities Legislation, Massachusetts Style,* 67 HARV. L. REV. 1349, 1349 (1954)).
87. *Id.*, *quoting* Leach, *supra* note 86, at 1349.
88. *Id.*
89. Wright v. Williams, 121 Cal. Rptr. 194, 199 n.2 (Ct. App. 1975).
90. Millwright v. Romer, 322 N.W.2d 30, 33 (Iowa 1982).
91. *See* Begleiter, *supra* note 1, at 249-52. Other areas in which malpractice actions have been brought are tax and estate planning and failure to research relevant law. *See id.* at 233-47. These areas are not relevant to an examination of the UPC and a description of these areas is therefore beyond the scope of this Article.
92. *E.g.*, Sorenson v. Fio Rito, 413 N.E.2d 47 (Ill. App. Ct. 1980); Cameron v. Montgomery, 225 N.W.2d 154 (Iowa 1975); *In re* Remsen, 415 N.Y.S.2d 370 (Sur. Ct. 1979).

area also involve delay in acting or failure to act. An attorney has been held liable for failure to file a wrongful death claim on behalf of decedent's sole heir, not advising the administratrix to list the claim as an estate asset and representing conflicting interests.[93] Failure to disclose to a beneficiary that the attorney represented other beneficiaries with adverse interests in a transaction involving the estate and a testamentary trust was held sufficient to form the basis of a malpractice action.[94] Failure to apply for a reduction in the administrator's bond, resulting in the estate's payment of higher premiums, has been held to support a malpractice action.[95] In addition the Alaska Supreme Court in *Linck v. Barokas & Martin*,[96] held the failure to advise a beneficiary to disclaim a portion of a bequest to save taxes was sufficient to state a cause of action for malpractice. In that case the defendant attorneys were developing an estate plan for decedent that was never effectuated because of his death. Under his will, his entire estate of $3 million passed to his wife. In such circumstances a disclaimer of a portion of the estate is often a valuable estate planning technique.[97] The defendant attorneys, however, failed to advise the widow to disclaim. The court ruled the plaintiffs stated a valid cause of action for professional negligence. The court further held damages included gift taxes the spouse paid in an attempt to minimize taxes, attorney and accountant fees, the children's loss of the use of the funds of the estate until their mother's death, and their loss of money used for payment of taxes and fees.[98]

---

93. Jenkins v. Wheeler, 316 S.E.2d 354, 358 (N.C. Ct. App. 1984) *rev. den.* 321 S.E.2d 136 (N.C. 1984).

94. Morales v. Field, DeGroff, Huppert & MacGowan, 160 Cal. Rptr. 239, 244 (Ct. App. 1979).

95. Geltman v. Levy, 207 N.Y.S.2d 366, 370-71 (App. Div. 1960) (cause sufficient to support malpractice suit if negligence attributable to the estate's attorney, rather than administrators of the estate).

96. 667 P.2d 171, 173-74 (Alaska 1983).

97. The receipt of all the assets by the spouse will result in no tax in the decedent's estate because of the marital deduction, I.R.C. § 2056 (1988), but the entire amount will be included in the surviving spouse's estate. I.R.C. § 2033 (1988). The decedent cannot use his unified credit, I.R.C. § 2010 (1988), in this situation. However, if the surviving spouse disclaims $600,000 of the decedent's estate, no tax will result in the decedent's estate and only $2,400,000 will be included in the surviving spouse's estate. Nor will any gift tax be payable because of the disclaimer. I.R.C. § 2518 (1988). However, I.R.C. § 2518(b)(2) requires the disclaimer be made within nine months of the decedent's death. In this case, an Alaska statute required the disclaimer be made within six months of the decedent's death. *See* ALASKA STAT. § 13.11.295 (1977), *quoted in* Linck v. Barokas & Martin, 667 P.2d 171, 172 n.1 (Alaska 1983).

98. *Linck*, 667 P.2d at 173-74. *See also* Kramer v. Belfi, 482 N.Y.S.2d 898 (App. Div. 1984), in which summary judgment against an attorney for failure to advise the executor of a husband's estate to renounce a trust created under the will of his deceased wife, with severe tax consequences to the husband's estate, was affirmed.

III. Provisions In New UPC Article II That Lessen The Likelihood Of Malpractice Litigation

In this section certain provisions of the new UPC that discourage malpractice litigation will be discussed. Before proceeding, it should be noted the focus of this article is on the malpractice implications of the Article II revisions. For this reason I may discuss a subsection rather than a complete section. In most cases the wisdom of the revisions is beyond the scope of this article. Scholars may disagree about the efficacy or desirability of some or all of the provisions discussed, but I will not enter any such debate here. Nor will I attempt to give a detailed explanation of the revised provision.[99] It is important to remember that one major method of preventing malpractice litigation is to provide a way for the problem to be corrected in the probate proceeding. Another is to provide a statutory solution responsive to the intent of most testators. As will be seen, most provisions of the new UPC that will be discussed employ one of these methods.

A. Section 2-101(b)—Negative Wills

On occasion a testator wants to exclude a person from taking property under the testator's will. Usually the testator desires that such a person take nothing from testator's estate, whether under the will or through intestacy. Under the majority American common law rule the only way to exclude an heir from taking by intestacy is to have a valid will disposing of all the testator's property.[100] Although the testator's wish could be accomplished it had to be done indirectly, and any slip prevented attaining the testator's objective. New subsection (b) was added to Section 2-101 of the UPC to reverse the majority rule and authorize the so-called "negative will." Though no malpractice case exactly on point has yet been reported, cases similar in nature have been litigated. In *Heyer v. Flaig*,[101] testatrix retained the defendant to prepare a will leaving all her property to her daughters. Although she was planning to marry, she wanted none of her estate (presumably including any intestate property) to go to her husband. Her husband, however, received part of the estate by statute.[102] The court upheld the malpractice action but disposed of the case on statute of limitations grounds.[103] It appears there is no

---

99. The comments to each section of Uniform Probate Code, art. II are particularly useful in this regard.

100. J. Andrew Heaton, Note, *The Intestate Claims of Heirs Excluded by Will: Should "Negative Wills" Be Enforced?*, 52 U. Chi. L. Rev. 177, 179 (1985).

101. 449 P.2d 161 (Cal. 1969).

102. *Id.* at 162.

103. The court held the statute of limitations began to run at the testator's death. *Id.* at 168.

significant difference between malpractice liability for failing to advise a testator of a post-testamentary spouse's rights and for failing to advise a testator of the possibility of an excluded person taking intestate.[104]

New Section 2-101 clearly solves this problem by permitting the negative will. If the client desires that a person take none of his estate by will or intestacy the will need only state that desire. The new UPC will give it effect. This rule should prevent any malpractice action based on a person who decedent desired would take nothing by intestacy.

B. Section 2-503—Dispensing Power

New Section 2-503 of the UPC enables a court to treat a document as complying with the formal requirements for execution, even if it does not do so, on proof by clear and convincing evidence that the decedent intended the document to be his or her will.[105] In other words, the court has the power to dispense with the formal requirements for execution based on clear testamentary intent. This section would cure most of the execution errors that have

---

104. *See* McAbee v. Edwards, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976); Walker v. Lawson, 526 N.E.2d 968 (Ind. 1988). Because these cases involved spouses, who could elect their statutory share, they are not exactly on point. These cases, however, are similar enough to demonstrate that under the majority American rule, a malpractice action could lie against an attorney for not arranging a client's affairs (by will, inter vivos gift or trust) to exclude a person the client wanted excluded, by merely excluding them in the will and not protecting against the person taking by intestacy.

105. UNIFORM PROBATE CODE § 2-503 (1990), which provides:

> Although a document or writing added upon a document was not executed in compliance with Section 2-502, the document or writing is treated as if it had been executed in compliance with that section if the proponent of the document or writing establishes by clear and convincing evidence that the decedent intended the document or writing to constitute (i) the decedent's will, (ii) a partial or complete revocation of the will, (iii) an addition to or an alteration of the will, or (iv) a partial or complete revival of his [or her] formerly revoked will or of a formerly revoked portion of the will.

The theory behind this section and use of it or a similar device in Australia, Israel and several Canadian provinces was brought to the attention of American lawyers by Professor John Langbein, now of Yale Law School, in a series of articles. *See* John H. Langbein, *Excusing Harmless Errors in the Execution of Wills: A Report on Australia's Tranquil Revolution in Probate Law*, 87 COLUM. L. REV. 1 (1987) [hereinafter Langbein, *Harmless Error*]; John H. Langbein, *Crumbling of the Wills Act: Australians Point the Way*, 65 A.B.A.J. 1192 (1979); John H. Langbein, *Substantial Compliance with the Wills Act*, 88 HARV. L. REV. 489 (1975) [hereinafter Langbein, *Substantial Compliance*].

generated the bulk of the malpractice actions over the past thirty years.[106]

It is also likely the section will cure a number of errors which, although they as of yet have generated no reported malpractice case, clearly should cause malpractice liability. The first such type of error is the "misplaced signature" case, which may arise in several different settings. If a state by statute requires a will be subscribed, and the will is signed in a place other than at the end, the entire will may be void or everything below the signature may be ignored.[107] A second example of this error occurs when the testator and witnesses sign a self-proving affidavit attached to the will but not the will itself. Many courts hold because such an affidavit is not a part of the will, the will is not signed and witnessed and is thus void.[108] Section 2-503 should solve such problems at the probate stage without the necessity of a malpractice action.[109]

It is also likely Section 2-503 could apply to the "crossed will" or *Pavlinko* case, named for a Pennsylvania case illustrating the error.[110] In this type of case mirror-image wills are prepared for two testators, almost invariably husband and wife. At the execution ceremony, through error, the husband signs the will prepared for the wife and the wife signs the will prepared for the husband. Several courts have held both wills invalid.[111] In a later, and perhaps more famous case,[112] the New York Court of Appeals admitted the will executed by the husband to probate and reformed it on the ground that "[u]nder such facts it would indeed be ironic—if not perverse—to state that because what has occurred is so obvious, and what was intended so clear, we must act to nullify rather than sustain this

---

106. *See, e.g.*, Licata v. Spector, 225 A.2d 28 (Conn. C.P. 1966) (will did not contain signatures of required number of witnesses); Price v. Holmes, 422 P.2d 976 (Kan. 1967) (will not executed in presence of attesting witnesses); Persche v. Jones, 387 N.W.2d 32 (S.D. 1986) (testator and witnesses did not sign in the presence of each other); Auric v. Continental Cas. Co., 331 N.W.2d 325 (Wis. 1983) (second witness did not sign will because of a mistake or confusion). *See* Begleiter, *supra* note 1, at 218-22 and Section II A, *supra*.

107. *E.g.*, N.Y. EST. POWERS & TRUSTS § 3-2.1 (McKinney 1981). Under New York Law, a bequest following the signature is void. *See In re* Estate of Mergenthaler, 474 N.Y.S.2d 253 (Sur. Ct. 1984).

108. *See, e.g.*, Boren v. Boren, 402 S.W.2d 728 (Tex. 1966); Estate of Ricketts, 773 P.2d 93 (Wash. Ct. App. 1989); Bruce H. Mann, *Self-Proving Affidavits and Formalism in Wills Adjudication,* 63 WASH. U.L.Q. 39 (1985). *See infra* Section III C.

109. In the first situation, a court might have to be concerned with the conflict between the statute requiring subscription and the dispensing power statute. Of course, if a state adopts the entire Article II, no such problem would occur since the UPC does not require subscription.

110. *In re* Estate of Pavlinko, 148 A.2d 528 (Pa. 1959).

111. *Id.* at 529-30.

112. *In re* Snide, 418 N.E.2d 656 (N.Y. 1981).

testamentary scheme."[113] Unfortunately, it was "obvious" and "clear" only to four members of the Court; a vigorous three-judge dissent emphasized the departure from precedent and the possible scope of the decision.[114]

It could be argued that whether Section 2-503 applies to this type of case depends on how the error is characterized. The error could be described as being that the testator left his will unsigned.[115] If it is so viewed, the error is clearly a violation of the signature requirement,[116] which Section 2-503 would cure. Alternatively, the error could be described as a lack of testamentary intent in that the testator never intended to execute the document he actually signed.[117] It could be argued that the dispensing power statute does not apply because testamentary intent is not an execution formality.

In all likelihood such an argument would fail. Testamentary intent is in fact included in the execution requirements of Section 2-502(c).[118] Therefore, the reference in Section 2-503 that the document "was not executed in compliance with Section 2-502" should apply.[119] In addition, testamentary intent may be considered as a formality similar to writing, signature and witnessing.[120] More importantly, the comments to Section 2-503 make it clear the section is intended to apply to this type of case,[121] and a jurisdiction that has had a dispensing power statute similar to Section 2-503 in force for some years has applied the statute to this type of case.[122] For all these reasons the switched will case should be governed by Section 2-503 and be solved at the probate stage, rendering a malpractice action unnecessary.

In *Guy v. Liederbach*[123] the testator, a Pennsylvania resident, hired a Pennsylvania attorney to draft a one-page will. Plaintiff, named as residuary beneficiary and executor, was one of the two

---

113. *Id.* at 657.

114. *See id.* at 658-59 (Jones, J., dissenting). The first part of this sentence should not be taken as a snide remark.

115. Langbein, *Harmless Error*, *supra* note 105, at 25. Some support for this view is found in Estate of Pavlinko, 148 A.2d 528, 529-30 (Pa. 1959).

116. UNIFORM PROBATE CODE § 2-502(a)(2) (1990).

117. *In re* Snide, 418 N.E.2d 656, 657 (N.Y. 1981); *In re* Estate of Pavlinko, 148 A.2d 528, 529 (Pa. 1959). Another view is that the testator executed a will containing mistaken provisions. *See* Langbein, *Harmless Error*, *supra* note 105, at 25; John H. Langbein & Lawrence W. Waggoner, *Reformation of Wills on the Ground of Mistake: Change of Direction in American Law?*, 130 U. PA. L. REV. 521, 562-66 (1982).

118. UNIFORM PROBATE CODE § 2-502(c) (1990) provides in relevant part: "Intent that the document constitute the testator's will can be established by extrinsic evidence . . . ."

119. UNIFORM PROBATE CODE § 2-503 (1990).

120. Langbein & Waggoner, *supra* note 117, at 565.

121. UNIFORM PROBATE CODE § 2-503 cmt. (1990).

122. Langbein, *Harmless Error*, *supra* note 105, at 24.

123. 459 A.2d 744 (Pa. 1983).

witnesses to the will. The testator owned property in New Jersey,[124] which had a statute invalidating bequests to necessary witnesses. Because the plaintiff was prohibited from inheriting testator's New Jersey property, she sued the attorney for malpractice. The court held a third party beneficiary could bring a malpractice action and remanded the case for a determination of whether plaintiff was in that category.[125]

It is possible, though not certain, that Section 2-503 would apply to this situation in states retaining the requirement that witnesses be competent or disinterested.[126] Two types of statutes exist.[127] The first type invalidates the will completely.[128] The second type, which was involved in *Guy v. Liederbach*,[129] upholds the will but either invalidates the bequest to the interested witness or reduces the bequest to the interested witness's intestate share.[130] Section 2-503 clearly applies to validate the will under the first type of statute. As to the second type it could be argued Section 2-503 has nothing to apply to because the will is not rendered invalid by the interested witness statute. Moreover, Section 2-503 says nothing about restoring the bequest to the interested witness. On the other hand there are indications that Section 2-503 may apply to validate the bequest to the interested witness in full, despite the fact that apparently no cases on this problem have arisen in jurisdictions that have enacted similar statutes.[131] Israel has adopted a dispensing power statute and has specified that the statute should apply to any defect in the capacity of the witnesses.[132] This indicates that at least one jurisdiction views the problem of interested witnesses as an execution requirement. Moreover, Professor John H. Langbein of Yale Law School, the person most responsible for bringing the dispensing power concept to the

---

124. The court assumed the testator was a New Jersey resident at his death, although this was not clear. *Id.* at 747 n.3.

125. *Id.* at 750-51.

126. It should be noted a malpractice action would not be necessary even under the old UPC. A beneficiary does not lose his or her bequest by being a witness to the will, nor is the will invalid for that reason. UNIFORM PROBATE CODE § 2-505 cmt. (1990).

127. *See* John B. Rees, Jr., *American Wills Statutes,* 46 VA. L. REV. 613, 625-34 (1960); Julian R. Kossow, *Probate Law and the Uniform Probate Code: "One for the Money. . . .",* 61 GEO. L.J. 1357, 1394-1400 (1973).

128. *Id.*

129. 459 A.2d 744 (Pa. 1983).

130. *See supra* note 127.

131. See Langbein, *Harmless Error, supra* note 105, for a discussion of the cases in these jurisdictions.

132. *See id.* at 48-51 (description of the Israeli statute). Although the statute speaks of validating the will, the specification of the capacity problem could arguably be read to imply that all aspects of the defect would be cured. *Id.*

attention of American lawyers and having it included in the UPC, has argued the dispensing power (or substantial compliance, a similar doctrine) should apply to validate a bequest to an interested witness. He suggests, as does the UPC,[133] that a gift to an interested witness should be viewed as a suspicious circumstance that could be challenged on the ground of undue influence:

> In jurisdictions which retain the competency requirement, the substantial compliance doctrine would have a similar effect, except that the burden of proof would be on those seeking to validate the will. The proponents [or in the case of an operative purging statute, the witness who stands to lose his legacy[134]] would have to show that the attesting witness who benefitted under the will had not in fact worked any imposition on the testator. When the witness takes only a token benefit, that fact alone should rebut the inference of imposition.[135]

It is at least possible that Section 2-503 could apply to solve the interested witness problem at the probate level.

C. Section 2-504—Self-Proved Will

The last subsection mentioned the type of misplaced signature case where the testator and attesting witnesses sign a self-proving affidavit but not the will and stated Section 2-503 should solve that problem. Actually, that section is not necessary because of an addition to Section 2-504 that authorizes self-proved wills. The new version of Section 2-504(c) states: "A signature affixed to a self-proving affidavit attached to a will is considered a signature affixed to the will, if necessary to prove the will's due execution."[136] In the absence of such an explicit provision, courts were split about the effect of a will signed only on a self-proving affidavit. Some cases held the will void,[137] while others held the affidavit was part of the will and thus validated the will.[138] The addition of Section 2-504(c) solves the problem and renders any malpractice action for such an error unnecessary.

D. Section 2-301—Premarital Will

In at least two cases testators about to be remarried have desired to leave their entire estate to their children to the exclusion of their

---

133. UNIFORM PROBATE CODE § 2-502 cmt. (1990).
134. Langbein, *Substantial Compliance*, *supra* note 105, at 516 n.105.
135. *Id.* at 516.
136. UNIFORM PROBATE CODE § 2-504(c) (1990).
137. *See, e.g., In re* Estate of Ricketts, 773 P.2d 93 (Wash. Ct. App. 1989); Boren v. Boren, 402 S.W.2d 728 (Tex. 1966).
138. *See, e.g., In re* Estate of Carter, 565 A.2d 933 (Del. 1989); Estate of Downie, No. 89-391, 1989 Iowa App. LEXIS 547 (Iowa Ct. App. Nov. 27, 1989).

intended spouse.[139] In both cases the state had a "pretermitted spouse" statute, awarding the spouse a portion of the estate in such cases. In each case, the attorneys ignored the statute, and the spouse took a portion of the estate.[140]

In a major revision to Section 2-301, the UPC has attempted to solve this problem. First, the new statute provides that any portion of the estate bequeathed to testator's children born before the marriage who are not children of the surviving spouse, or that passes to the issue of such child under the antilapse statute[141] or through the residuary estate as a result of a failed bequest,[142] does not pass to the pretermitted spouse.[143] Second, the statute provides the spouse will not take under the statute:

> 1. If the will or other evidence shows the will was made in contemplation of the marriage;[144]
> 2. The will expresses it is to be effective despite the marriage;[145] or
> 3. The testator transfers assets outside the will to the spouse, and the testator's intent, as shown by his or her statements or inferred from the amount of the transfer or other evidence, is that the transfer be in place of a bequest in the will.[146]

In addition, the spouse in this situation cannot take both an intestate share as a pretermitted spouse and an elective share if the total is greater than the elective share.[147] The share under Section 2-301 counts toward making up the elective share.[148] Most malpractice cases of this type will come under one of the situations covered by the statute; malpractice litigation rarely will be necessary.

E. Section 2-606—Ademption by Extinction

Under the common law doctrine of ademption by extinction, a specific devise or bequest is voided if the property bequeathed is not a part of testator's estate at death. Surprisingly, two recent malpractice cases have involved this problem. In one case[149] a member of a law firm prepared a will for a testator leaving all the decedent's

---

139. Heyer v. Flaig, 449 P.2d 161 (Cal. 1969); McAbee v. Edwards, 340 So. 2d 1167 (Fla. Dist. Ct. App. 1976).

140. *Heyer*, 449 P.2d at 163; *McAbee*, 340 So. 2d at 1168.

141. UNIFORM PROBATE CODE § 2-603 (1990).

142. UNIFORM PROBATE CODE § 2-604 (1990).

143. UNIFORM PROBATE CODE § 2-301(a) (1990). This would have prevented the spouses from taking anything (except an elective share, if applicable) in each of the cases cited *supra* in note 139.

144. UNIFORM PROBATE CODE § 2-301(a)(1) (1990).

145. UNIFORM PROBATE CODE § 2-301(a)(2) (1990).

146. UNIFORM PROBATE CODE § 2-301(a)(3) (1990).

147. UNIFORM PROBATE CODE § 2-301 cmt. (1990).

148. *Id.*

149. Stangland v. Brock, 747 P.2d 464 (Wash. 1987).

real property (primarily farmland) to Stangland and Kintschi. About three years later another attorney for the same firm prepared a real estate contract selling the farmland. The second attorney neither reviewed the will file nor notified the attorney who drafted testator's will. When the testator died three months later, a dispute arose about whether the contract converted the real estate to personalty, which would pass under the residuary clause in testator's will. If so, the bequest to the plaintiffs would adeem. The probate action was settled and plaintiffs sued both attorneys and the law firm. The court held both attorneys owed a duty to plaintiffs, but that the attorneys acted in accordance with the applicable standard of care.[150] The drafting attorney performed his task competently; he had no notice of and no reason to anticipate the sale of the real estate. Absent such notice, no reasonable lawyer would be expected to foresee and draft against the events that occurred three years later.[151] As to the allegation that the drafting attorney should have advised the client of the effect of the contract on the will (which he should have known about because the firm's lawyers received memos on new matters), the court held the imposition of such a duty would expand the obligation of lawyers beyond reasonable limits. No continuing obligation exists on a will drafter to monitor the testator's management of his property to ensure the estate plan is maintained. The imposition of such a duty would prevent an attorney from providing clients with reliable services economically and would be an overwhelming burden on an attorney's practice.[152] The court further held it would be unreasonable to impute the knowledge of the contents of decedent's will to the attorney drafting the real estate contract. Therefore, that attorney had no duty to inform testator of the effect of the contract on the will.[153] A vigorous dissent argued decedent had a broader agreement with the law firm as a whole to provide legal services, and that in modern times, individuals must rely on the firm as a whole to look after their interests, particularly when the client has no choice as to who does the work.[154] The client has a right to rely on the firm as a whole to effectuate his best interests.[155] Part of the firm's duty is communication among lawyers concerning individual transactions of a client and consideration of the effect of each such transaction on the client's general interests.[156]

In a factually similar Iowa case[157] the decedent's will devised to the beneficiary one-half her interest in real property bequeathed to

150. *Id.* at 468-70.
151. *Id.* at 469-70.
152. *Id.*
153. *Id.* at 470.
154. *Stangland*, 747 P.2d at 471-72 (Goodloe, J., dissenting).
155. *Id.* at 472.
156. *Id.* at 472-73.
157. Schreiner v. Scoville, 410 N.W.2d 679 (Iowa 1987).

testatrix under the will of another person who recently had died and whose estate was not yet distributed.[158] Seven months later the attorney who drafted the original will drafted and witnessed a codicil expressly reaffirming the bequest.[159] Less than one month after the execution of the codicil, the attorney brought an action for partition by sale of the real property.[160] Eight months later the parcel was sold and testatrix received cash for her interest. Testatrix died nine months later.[161] In the probate proceeding the court determined the devise of land had been adeemed by the sale and held the proceeds of sale passed under the residuary clause of the will.[162] The beneficiary sued the attorney for malpractice, alleging that the testatrix's true intent that he receive the proceeds of sale was not effectuated, and that the attorney negligently failed to advise the testatrix of the effect of the partition sale on her will. The court, in a rather confusing opinion turning on privity, held a cause of action had been stated.[163]

New Section 2-606 of the UPC attempts to resolve this problem in the probate proceeding. The new section provides that the specific devisee has the right to real or tangible personal property owned by testator at death and acquired as a replacement for specifically devised real or tangible personal property.[164] Moreover, the beneficiary receives the value of adeemed property unless the facts and circumstances indicate the testator intended ademption to occur, or ademption is consistent with testator's distribution plan.[165] This creates a "mild

---

158. *Id.* at 680. The will also devised to the beneficiary a one-half interest in the residue of the estate. *Id.*

159. *Id.* The codicil also removed the beneficiary's one-half interest in the residue of the estate, however. *Id.*

160. *Id.*

161. *Id.* at 680.

162. *Id.* The probate court found that the ademption occurred when the sale changed the testatrix's interest in real property to one in personal property (cash), and because the testatrix had made no express bequest of the personal property, it passed to the four nephews named in the residuary clause of the codicil. *Id.*

163. *Id.* at 683. The opinion is confusing because the court held the complaint stated a cause of action and did not require privity, but it also held that a malpractice action is allowable "only when as a direct result of the lawyer's professional negligence the testator's intent *as expressed in the testamentary instruments* is frustrated in whole or in part and the beneficiary's interest in the estate is either lost, diminished, or unrealized." *Id.* at 683 (emphasis added). There is no way the complaint could satisfy this test. The only intent shown by the will was that the beneficiary receive testatrix's interest in the real estate. However, the court appeared to create an exception to its own rule, holding that at a trial the evidence might show that the will, the codicil, and the partition were interrelated, closely connected in time, and that the attorney represented the testatrix in all three transactions. *Id.* at 684. If so, the transactions became linked, creating a duty on the attorney to inform testatrix of the effect of the sale on the will. *See* Begleiter, *supra* note 1, at 260-63 (discussing the unclear scope of the exception created in *Schreiner*).

164. UNIFORM PROBATE CODE § 2-606(a)(5) (1990).

165. UNIFORM PROBATE CODE § 2-606(a)(6) (1990). The section provides that

presumption against ademption by extinction."[166] The adoption of the "intent" theory of ademption would solve the problem in both cases discussed in this subsection because the facts and circumstances, to the extent stated in the malpractice opinions, do not indicate any intent that the bequests be adeemed. Moreover, under new Section 2-606, the probate proceeding would solve these cases without any malpractice action.

F. Section 2-702—Survival

Husband's will provides that he bequeaths all his property to his wife if she survives him by thirty days. The will further provides that if his wife and he die in or from a common disaster or if his wife predeceases him, his estate should be divided equally among his nephews. Wife's will contained mirror image provisions. Husband dies suddenly of a stroke. Four days later, wife dies after a long battle with cancer.[167] Because the husband's will contained no provision covering this situation, the husband's estate passes intestate.[168] The nephews sue the drafting attorneys for malpractice, contending the intentions of the decedents were that the property of both was to go to the nephews if neither testator survived the other by thirty days, but that defendant attorneys negligently failed to include such a provision in the will.[169] The Illinois Supreme Court held a complaint based on similar facts stated a valid cause of action for legal malpractice.[170]

Many wills contain common disaster clauses. In a case such as that set forth above, however, the clause does not apply because testators do not die as a result of the same event. Moreover, common

---

the beneficiary has a right to:

> unless the facts and circumstances indicate that ademption of the devise was intended by the testator or ademption of the devise is consistent with the testator's manifested plan of distribution, the value of the specifically devised property to the extent the specifically devised property is not in the testator's estate at death and its value or its replacement is not covered by paragraphs (1) through (5).

Id.

166. UNIFORM PROBATE CODE § 2-606 cmt. (1990).

167. These facts are based on Ogle v. Fuiten, 466 N.E.2d 224 (Ill. 1984), except that the provision bequeathing the estate to the nephews if the wife predeceased the husband was not contained in the will involved in that case and that the wife died fifteen days after the husband. I have altered the facts to bring the case within UNIFORM PROBATE CODE § 2-702 (1990). These minor alterations, however, do not affect the holding of the case that the failure to include the provisions desired supported a cause of action for malpractice.

168. *In re* Estate of Smith, 385 N.E.2d 363 (Ill. App. Ct. 1979) (discussed in Ogle v. Fuiten, 466 N.E.2d 224 (Ill. 1984).

169. Ogle v. Fuiten, 466 N.E.2d 224 (Ill. 1984).

170. *Id.* at 227.

disaster clauses often contain a statement that the clause only applies if it cannot be determined whether the beneficiary or the testator survived. Such a clause does not cover the case where the beneficiary survives the testator but then dies a short time later. The UPC attempts to solve these problems in new Section 2-702, which applies not only to wills but to inter vivos trusts.[171] The section provides that, except as provided in the section,[172] for the purposes of the UPC and for a donative provision of a governing instrument, an individual "who is not established by clear and convincing evidence to have survived an event, including the death of another individual, by 120 hours is deemed to have predeceased the event."[173] The section does not apply "if the governing instrument contains language dealing explicitly with simultaneous deaths or deaths in a common disaster *and that language is operable under the facts of the case.*"[174] If this statute had been applicable to the facts posited at the beginning of this subsection,[175] the wife would have been deemed to have predeceased her husband because even though the will contained language dealing with a common disaster, the common disaster language would be inapplicable to the facts because the husband and wife did not die as a result of a common disaster. Therefore, the section would mandate that the wife be deemed to predecease her husband because she did not survive him by 120 hours, and the estate would pass to the nephews under the will provisions. Section 2-702 solves this problem without resort to malpractice litigation. The section will no doubt also be useful in solving problems in situations where persons die as a result of a common accident, although it is clear which person survived. Thus, new Section 2-702 will decrease the necessity of malpractice in some but not all[176] survival cases.

---

171. The constructional rules of Article II, Part 7 apply to a "governing instrument," which is a defined term. Section 1-201(19) defines governing instrument as

> a deed, will, trust, insurance or annuity policy, account with POD designation, security registered in beneficiary form (TOD), pension, profit-sharing, retirement, or similar benefit plan, instrument creating or exercising a power of appointment or a power of attorney, or a donative, appointive, or nominative instrument of any other type.

UNIFORM PROBATE CODE § 1-201(19) (1990).

172. TOD security registrations are excepted. UNIFORM PROBATE CODE § 2-702(a), (b) (1990). Section 2-702(d) of the UPC contains four other exceptions, none of which are applicable to this discussion. *See* UNIFORM PROBATE CODE § 2-707(d) (1990).

173. UNIFORM PROBATE CODE § 2-702(a), (b) (1990).

174. UNIFORM PROBATE CODE § 2-702(d)(1) (1990) (emphasis added).

175. *See supra* notes 145-47 and accompanying text.

176. For example, Section 2-702 would not have applied to the actual facts of Ogle v. Fuiten, 466 N.E.2d 224 (Ill. 1984) because the wife survived the husband by fifteen days.

G. Section 2-804—Divorce

A marries B. Each has two children by a previous marriage. In their wills, each leaves his or her property to the other spouse or, if the other spouse predeceases, to the testator's children and to the spouse's children. A and B divorce. A dies. In most states it is clear by statute or decision that divorce revokes the bequest to B.[177] The question is whether B's children take. The answer in most states appears to be that they do.[178] Although each situation is different, in at least some if not most cases it is likely that A did not desire this result. Though no malpractice action based on these facts has been discovered, it is not hard to imagine a relative of A bringing such an action based on the attorney's failure to advise A the divorce would not revoke the bequest to the former spouse's relatives. Nor would it be unlikely that such an action would be successful. The UPC has been revised to solve this problem; a divorce revokes not only bequests to the former spouse but also bequests to relatives of the former spouse.[179] In addition, several cases have considered the question whether a statute governing the effect of divorce on wills applies to will substitutes, such as life insurance policies or revocable trusts.[180] Revised Section 2-804 makes its provisions broadly applicable to nontestamentary instruments executed by the divorced individual before the divorce.[181] In addition, the revocatory effect of divorce extends to powers of appointment and nominations in any fiduciary or representative capacity.[182] Any other provisions of the governing instrument not revoked by the divorce are given effect as if the former spouse and his relatives disclaimed the revoked provision or, in the case of nominations, as if they predeceased the testator or grantor.[183] The comments are correct in touting Section 2-804 as "the most comprehensive provision of its kind."[184] Any malpractice problems based on failure to advise the testator on the effect of the divorce or on relatives of the former spouse receiving property should

---

177. WILLIAM M. MCGOVERN, SHELDON F. KURTZ & JAN ELLEN REIN, WILLS, TRUSTS AND ESTATES 221 (1988). *See* UNIFORM PROBATE CODE § 2-508 (1975) (prior version).

178. MCGOVERN, ET AL., *supra* note 177, at 221. *See, e.g.*, Porter v. Porter, 286 N.W.2d 649 (Iowa 1979); Clymer v. Mayo, 473 N.E.2d 1084 (Mass. 1985); *In re* Estate of Coffed, 387 N.E.2d 1209 (N.Y. 1979); Miller v. First Nat. Bank & Trust Co., 637 P.2d 75 (Okla. 1981).

179. UNIFORM PROBATE CODE § 2-804(b)(1) (1990). "Relative of the divorced individual's former spouse" is defined in UNIFORM PROBATE CODE § 2-804(a)(5) (1990).

180. *Clymer*, 473 N.E.2d at 1093; *Miller*, 637 P.2d at 77.

181. UNIFORM PROBATE CODE § 2-804(a)(4) (1990).

182. UNIFORM PROBATE CODE § 2-804(b)(1) (1990).

183. UNIFORM PROBATE CODE § 2-804(d) (1990).

184. UNIFORM PROBATE CODE § 2-804 cmt. (1990).

be ameliorated without the necessity of a malpractice action.

H. Part 7 - Extension of Constructional Rules to Other Dispositive Instruments

The importance of nonprobate transfers in estate planning is increasing.[185] Malpractice actions involving nonprobate transfers have been brought.[186] Attorneys will tend to become more familiar with constructional rules if the same rules are made applicable to both wills and will substitutes, as far as possible.[187] This is partly because the rules will be encountered with increasing frequency and partly because the more experience an attorney has with the rules, the better he will tend to remember, understand, and use the rules competently. The increase in competence caused by this process should result in less errors and less malpractice actions. The revision of Article II explicitly recognized this[188] by enacting Part 7, which provides constructional rules applicable to both wills and will substitutes.[189] The extension of constructional rules to will substitutes should simplify the law and "bring the law of probate and nonprobate transfers into greater unison,"[190] in addition to decreasing malpractice litigation.

IV. Provisions In New UPC Article II That Increase The Likelihood Of Malpractice Litigation

If a beneficiary dies before the testator, his or her gift lapses.[191] For many years legislatures have passed statutes to prevent the lapse of at least some testamentary bequests. Almost all states have somewhat similar statutes to prevent lapse in certain cases, although the details differ.[192] A majority of states do not apply the antilapse statute if a contrary intent appears in the will.[193] A bequest "To A

---

185. *See* John H. Langbein, *The Twentieth-Century Revolution in Family Wealth Transmission*, 86 Mich. L. Rev. 722 (1988); John H. Langbein, *The Nonprobate Revolution and the Future of the Law of Succession*, 97 Harv. L. Rev. 1108 (1984); Uniform Probate Code Article II, prefatory note (1990).

186. Krawczyk v. Stingle, 543 A.2d 733 (Conn. 1988) (inter vivos trust); McLane v. Russell, 512 N.E.2d 366 (Ill. App. Ct. 1987), *aff'd*, 546 N.E.2d 499 (Ill. 1989) (joint tenancy); Bormaster v. Baldridge, 723 S.W.2d 533 (Mo. Ct. App. 1987) (amendment to inter vivos trust).

187. Obviously some differences may be necessary depending on the type of instrument and form of property (*e.g.*, joint tenancy, life insurance) involved.

188. Uniform Probate Code Article II, prefatory note (1990).

189. Uniform Probate Code §§ 2-701, 2-711 (1990).

190. Uniform Probate Code Article II, prefatory note (1990).

191. Thomas E. Atkinson, Handbook of the law of Wills 777 (2d ed. 1953).

192. *Id.* at 779.

193. *Id.* at 780.

if A survives me" has, by most of the courts deciding the issue, been held to show a sufficient contrary intent to defeat the antilapse statute.[194] Because antilapse statutes have existed for a long period of time and the case law holding that "if he survives me" evidences a contrary intent under the antilapse statute spans many years,[195] it is reasonable to assume most lawyers are familiar with the rule and that some lawyers have incorporated the "if he survives me" language into their will forms. Lawyers do this believing it will negate the operation of the antilapse statute.

One of the surest ways to generate malpractice litigation is to change the effect of language that attorneys are comfortable with and have used for many years. This is especially true if some of the attorneys using the language do not do a significant amount of estate planning. Such attorneys are less likely to notice new developments in wills law and more likely to rely on old forms than are lawyers who have a significant estate planning practice. This is exactly what the revision of Article II has accomplished. The revision provides that the words "if he survives me" will not constitute a contrary intent and will not negate the operation of the antilapse statute.[196] Thus, if the revised provisions are adopted, property bequeathed "to A if he survives me" will pass to A's descendants who survive the testator, rather than to the residuary beneficiaries under A's will, if A is a relative of testator within the class designated in the statute.[197]

---

194. *Id.* at 780-81.

195. *Id.* at 779-80.

196. UNIFORM PROBATE CODE §§ 2-603(b)(3), 2-706(b)(3) (1990). Section 2-603(b)(3) provides: "For the purposes of Section 2-601, words of survivorship, such as in a devise to an individual 'if he survives me,' or in a devise to 'my surviving children,' are not, in the absence of additional evidence, a sufficient indication of an intent contrary to the application of this section." Section 2-706 (b)(3), applicable to wills and testamentary substitutes, is similar. Section 2-601 provides that the rules of Part 6 control the construction of a will in the absence of a finding of a contrary intention.

Sections 2-603 (wills) and 2-706 (testamentary substitutes) are the antilapse provisions of revised Article II of the UPC. A detailed explanation of these comprehensive and complex provisions is beyond the scope of this article. Basically, the statute saves dispositions to a grandparent, a descendant of a grandparent and a stepchild for the descendants of the beneficiary if the beneficiary predeceases the testator or donor. The statute applies to class gifts with certain exceptions and applies to gifts to persons who are dead at the time the will is executed (so-called "void" gifts at common law). The statute is negated by an alternate bequest in most cases. The statute also provides rules covering the priority of substitute gifts.

197. *See* UNIFORM PROBATE CODE §§ 2-603, 2-706 (1990). This is not to say that as a policy matter this is not a good rule. It is clear most experienced wills lawyers and scholars would prefer a drafter to say "To A, but if he shall not survive me, to B" rather than "To A if he survives me." The former reveals testator's intent far more clearly than the latter and expressly provides a substitute taker. That the former is a better method of drafting is one thing. To require that form be used to negate the antilapse statute is quite another matter.

The adoption of such a rule, which is contrary to the long established law of most states and flies in the face of the usage of the "if he survives me" language in many will forms of practicing lawyers, is likely to cause an increase in malpractice litigation.

The comment to Section 2-603 recognizes the problem but denies it is a problem in the same breath. The comment notes if evidence shows the lawyer and client discussed the question, and the client determined that he did not wish the antilapse statute to apply:

> then the combination of the words of surivivorship and the extrinsic evidence of the client's intention would support a finding of a contrary intention under Section 2-601. . . . For this reason, Sections 2-601 and 2-603 will not expose lawyers to malpractice liability for the amount that, in the absence of the finding of the contrary intention, would have passed under the anti-lapse statute to a deceased devisee's descendants. The success of a malpractice claim depends upon sufficient evidence of a client's intention and the lawyer's failure to carry out that intention. In a case in which there is evidence that the client did not want the antilapse statute to apply, that evidence would support a finding of a contrary intention under Section 2-601, thus preventing the client's intention from being defeated by Section 2-603 and protecting the lawyer from liability . . . .[198]

This explanation, when evaluated closely, is inadequate. Moreover, at least in one sense, it does not address the true problem.

Perhaps the most important factor the comment overlooks is that it is not only malpractice liability that hurts the lawyer, but the mere commencement of the malpractice action. The lawyer suffers unfavorable publicity, with the possible loss of clients and other effects on his practice, and a possible increase in malpractice insurance premiums. That the lawyer might ultimately win the malpractice action is helpful but will not repair the damage that commencement of the action causes. The comment is only concerned with liability and ignores the harm engendered by the commencement of malpractice actions.

In addition, the conclusion of the comment that a contrary intent easily can be shown by extrinsic evidence is by no means certain. First, while it is true that the revision of Section 2-601 appears to show an intent to liberalize the admission of extrinsic evidence in the interpretation of wills, this is explained in a comment and is not indicated in the text of Section 2-601.[199] It is surely possible, indeed likely, that a court will give little weight to the comment in the face

---

198. UNIFORM PROBATE CODE § 2-603 cmt. (1990).

199. UNIFORM PROBATE CODE § 2-601 cmt. (1990). Section 2-601 provides only that "[i]n the absence of a finding of a contrary intention, the rules of construction in this Part control the construction of a will."

of the often repeated rule that the interpretation of a will is governed by the language of the will taken as a whole and the circumstances surrounding the testator at the time of execution.[200] It is probable most courts will be reluctant to depart from this long standing rule on will interpretation to the extent envisioned in the comment.

Even if this obstacle is overcome, a showing of contrary intent is not certain. The comment to Section 2-603 indicates that a memorandum or pre-execution letter from the attorney to the testator stating that the attorney used the word "surviving" to negate the statute would be evidence tending to support a finding of contrary intent.[201] The existence of such a memorandum would be rare. None of the malpractice cases so far brought show any evidence of such writings.[202] Indeed, it is far more likely that any evidence of the attorney's meaning in using survivorship language or the testator's desires with respect to the question of lapse will be totally oral. Whether such evidence is admissible in interpreting the will is problematical, especially in light of the rule that direct declarations of the testator's intent generally are not admissible, except to resolve an equivocation.[203] Evidence of the oral contract for the drafting of the will and its terms is admissible in the malpractice action, however.[204]

Lastly, in many cases the issue of whether the testator had a contrary intent does not arise in the probate proceeding. The reason is that many probate proceedings, including ones in which malpractice claims are brought, are settled.[205] Part of the reason for this may be that family members are involved in the probate proceeding and the disappointed beneficiary would rather bring an action against the drafting attorney than generate family controversy. Whatever the reason, in some cases the question of contrary intent will never arise in the probate action.

It is very clear that by reversing the settled law that the language "if he survives me" shows an intent to negate the antilapse statute, Sections 2-603(b)(3) and 2-706(b)(3) create the conditions for an increase in malpractice actions. The explanation given by the UPC

---

200. THOMAS E. ATKINSON, HANDBOOK OF THE LAW OF WILLS 810-11 (2d ed. 1953).

201. UNIFORM PROBATE CODE § 2-603 cmt. (1990).

202. *See* cases discussed in Begleiter, *supra* note 1.

203. THOMAS E. ATKINSON, HANDBOOK OF THE LAW OF WILLS 810 (2d ed. 1953).

204. Begleiter, *supra* note 1, at 201 and n.46.

205. *E.g.*, Ventura County Humane Soc'y for Prevention of Cruelty to Children & Animals v. Holloway, 115 Cal. Rptr. 464 (Ct. App. 1974); Stowe v. Smith, 441 A.2d 81 (Conn. 1981); McAbee v. Edwards, 340 So. 2d 1167 (Fla. Ct. App. 1976); Persche v. Jones, 387 N.W.2d 32 (S.D. 1986); Stangland v. Brock, 747 P.2d 464 (Wash. 1987).

comment does not adequately address the problem. It is highly likely that in many cases the issue will not be solved by will interpretation, but in a malpractice action.

V. CONCLUSION

Revising Article II of the Uniform Probate Code has produced many improvements in the UPC. The UPC has clarified the law in many areas.[206] Many new and important provisions recently have been added to the UPC, some of which were discussed in this article. The drafters made improvements in other sections as well. While arguments may arise about the policy basis of individual sections, the revisions of Article II clearly mark a significant point in the development of the UPC.

Many of the new and revised provisions solve problems that, if not addressed, could result in the initiation of malpractice actions. Indeed, with the exception of one provision in the antilapse sections,[207] the provisions of the revised Article II examined in this article uniformly lessen the chances of malpractice litigation ensuing. For this, the drafters are to be commended.

206. *E.g.*, UNIFORM PROBATE CODE §§ 2-507, 2-509 (1990) (revocation and revival).

207. UNIFORM PROBATE CODE §§ 2-603(b)(3), 2-706(b)(3) (1990).

DUELING NARRATIVES IN *AN AMERICAN TRAGEDY* AND THE CRIMINAL LAW

VANESSA LAIRD*

The assertion of connections between law and literature has become commonplace, so commonplace, in fact, as to have provoked a critical backlash.[1] Two of these connections constitute my starting point: the fact that the narrative form is characteristic of both legal and literary discourse[2] and the fact that law and literature are social products, shaped by and shaping the ideological[3] context in which they are produced.[4] I will pursue the argument that one of the ways law and literature lead us to believe that it is natural to make sense of events in one way rather than another is through their common

---

* B.A. Williams College; M.A. University of Oxford; J.D. Stanford University. Lecturer (assistant professor) in Law, Faculty of Law, University of Sheffield, Sheffield, England. The author would like to thank Roger Brownsword, David Laird, and Robert Weisberg for their helpful comments on earlier drafts of this piece.

1. *See, e.g.*, Robert Weisberg, *The Law-Literature Enterprise*, 1 YALE J.L. & HUMAN. 1 (1988); Robin L. West, *Adjudication Is Not Interpretation: Some Reservations About the Law-as-Literature Movement*, 54 TENN. L. REV. 203 (1987).

2. The prevalence of stories in the legal world has been widely, perhaps almost excessively, noted. *See, e.g.*, DAVID R. PAPKE, NARRATIVE AND THE LEGAL DISCOURSE 1 (1990); James R. Elkins, *From the Symposium Editor*, 40 J. LEGAL EDUC. 1 (1990); Kim L. Scheppele, *Foreword: Telling Stories*, 87 MICH. L. REV. 2073 (1989). Examples of writings about particular forms of legal stories include Douglas W. Maynard, *Narratives and Narrative Structure in Plea Bargaining*, 22 LAW & SOC. REV. 449 (1988); William M. O'Barr and John M. Conley, *Litigant Satisfaction Versus Legal Adequacy in Small Claims Court Narratives*, 19 LAW & SOC. REV. 661 (1985); David R. Papke, *Discharge as Denouement: Appreciating the Storytelling of Appellate Opinions*, 40 J. LEGAL EDUC. 145 (1990); Kathryn H. Snedaker, *Storytelling in Opening Statements: Framing the Argument of the Trial*, 10 AM. J. TRIAL ADVOC. 15 (1986). Some useful analyses or characterizations of law based on the constitutive role played by narrative appear in BERNARD S. JACKSON, LAW, FACT, AND NARRATIVE COHERENCE 10-11 (1988); JAMES B. WHITE, HERACLES' BOW 165-75 (1985); Bernard S. Jackson, *Narrative Theories and Legal Discourse*, *in* NARRATIVE IN CULTURE 23-50 (Cristopher Nash ed., 1990). A more comprehensive listing of the literature relating to narrative aspects of legal discourse is found in James R. Elkins, *A Bibliography of Narrative*, 40 J. LEGAL EDUC. 203 (1990).

3. The term "ideological" is used in a myriad of ways. *See* TERRY EAGLETON, IDEOLOGY 1-2 (1991). I use it here in one of the senses Eagleton describes to refer to "the medium in which conscious social actors make sense of their world." *Id.* at 2.

4. Brook Thomas, among others, has noted that legal and literary narratives reflect and respond to the conditions of their production and reception. *See* BROOK THOMAS, CROSS-EXAMINATIONS OF LAW AND LITERATURE 6-7 (1987).

deployment of the narritive form.[5] My argument will be embedded in an examination of the novel *An American Tragedy* by Theodore Dreiser.

I hope to keep the law-literature skeptics at bay by showing that study of *An American Tragedy* assists an understanding of three aspects of legal doctrine: the fact-suppressing and therefore ideological character of legal narrative; the criminal law's bias toward intentionalist stories;[6] and its corresponding attempt to neutralize the threat determinism—evoked as a theory either of all human action or of the actions of an unexceptional individual[7]—poses to the

---

5. Although the narrative representations of human behavior that appear in legal contexts tend to present themselves as ideologically neutral depictions of a found world, contemporary writings in law and the humanities have disputed both the found character and the neutrality of narrative. *See, e.g.*, JACKSON, LAW, FACT, AND NARRATIVE COHERENCE, *supra* note 2, at 10; HAYDEN WHITE, TROPICS OF DISCOURSE 82-85 (1978); WHITE, *supra* note 2, at 175; Scheppele, *supra* note 2, at 2075, 2090. The two points are intertwined, though their relationship is generally left implicit. An argument in which the two points figure might be sketched roughly as follows. Telling a story requires criteria for sorting out the relevant from the irrelevant facts and for deciding upon the language in which to describe these facts. *See* WHITE, *supra* note 2, at 175 ("[N]o story can include everything . . . every story is a reduction, a fiction, made from a certain point of view."). These criteria come not from the facts themselves but from the the theories or ideas about the way the world works that are available to us. *See* JACKSON, LAW, FACT, AND NARRATIVE COHERENCE, *supra* note 2, at 10-11 ("[I]nference from one fact to another . . . involves a relationship of plausibility. And plausibility is constructed . . . in terms of narrative models which in their structure may be universal but in their content are socially and culturally contingent, models which reflect both common experience (at least what is socially constructed and common as common experience) and the social and cultural values that inform such collective representations."); Steven L. Winter, *The Cognitive Dimension of the* Agon *Between Legal Power and Narrative Meaning*, 87 MICH. L. REV. 2225, 2228 (1989) ("In narrative, we take experience and configure it in a conventional and comprehensible form."). Any given story will thus, by defining for the reader what is relevant, privilege the theories or ideologies from which its judgments of relevance derive. *See* Papke, *Discharge as Denouement*, *supra* note 2, at 158 ("Narratives are never merely descriptive, or fanciful, they are also explanatory. Storytellers, whether individuals, business or government institutions, select characters and events, place the events in sequence, and imply that the sequencing is normal, comprehensible, and desirable. Stories, as a result, establish a complex normative environment.").

6. *See* H.L.A. HART, PUNISHMENT AND RESPONSIBILITY 28 (1968); MARK KELMAN, A GUIDE TO CRITICAL LEGAL STUDIES 89 (1987) [hereinafter GUIDE]; Mark Kelman, *Interpretive Construction in the Substantive Criminal Law*, 33 STAN. L. REV. 591, 598 (1981).

7. For a discussion of the intentionalist and determinist models of action, see *infra* text accompanying notes 11-18. Mark Kelman has argued that we are simultaneously drawn both to intentionalist and determinist expression and that legal discourse suppresses conflict between them by creating the impression that we can "*descriptively* identify domains of freedom and distinguish them from domains of choicelessness . . . ." KELMAN, GUIDE, *supra* note 6, at 87 (emphasis in original).

justifiability of that bias. *An American Tragedy* illustrates the constructed character of narrative particularly well because it juxtaposes different accounts of the same crime.[8] The novel provides three different narrative explanations of Clyde Griffiths' role in the death of his girlfriend, Roberta Alden: the novel's own linear narrative ("Dreiser's narrative"),[9] which describes Clyde's life from the age of twelve until his execution for Roberta's murder, and the accounts

---

Because I can be taken to be offering a literary elaboration of Kelman's argument, it is worth responding at the outset to two criticisms that have been leveled against it.

First, John Stick and Ken Kress have accused Kelman of confusing description of a particular act as involuntary with a commitment to determinist discourse in general. *See* John Stick, *Charting the Development of Critical Legal Studies*, 88 COLUM. L. REV. 407, 414 (1988); Ken Kress, *Legal Indeterminacy*, 77 CAL. L. REV. 283, 316 (1989). They are right to point out that an involuntary act does not establish the impossibility of free will. For us to be comfortable with the practice of blaming that has been institutionalized in the criminal law, however, we must believe free will to be the norm and not a mere possibility. To the extent that, as Kelman argues, our propensity for determinist description cannot be confined to exceptional cases, it will tend to erode our satisfaction with the intentionalist model and with the criminal law.

Second, Kress has criticized Kelman implicitly for simply accepting "the popular notion that determinism is incompatible with free will" and ignoring "most current philosophers' preference" for the compatibalist position. Kress, *supra* this note, at 316. The compatibalist stance, as traditionally conceived, reconciles free will with determinism by defining it as voluntariness rather than as voluntariness plus origination. TED HONDERICH, A THEORY OF DETERMINISM; 487 (1988) Contrary to Kress' suggestion, this solution, which is unsatisfactory in light of our anxiety about assigning moral responsibility when voluntariness is present without origination, *see* discussion *infra* note 14, continues to generate philosophical controversy. *See, e.g.*, HONDERICH, *supra* this note; MARTHA KLEIN, DETERMINISM, BLAMEWORTHINESS AND DEPRIVATION (1990).

8. James Boyd White has noted that competition between or among stories can help to reveal the constructed character of narrative. *See* WHITE, *supra* note 2, at 175 ("In looking at competing stories, and trying to decide between them, . . . we thus naturally think in terms of inclusion and exclusion . . . ."). And Richard Delgado has deployed the technique of opposing different tellings of the same event in order to illuminate their character and implications. Richard Delgado, *Storytelling for Oppositionists and Others: A Plea for Narrative*, 87 MICH. L. REV. 2411, 2418-26 (1989) (undermining the authority of a "stock story" about the rejection of a black lawyer for a law school teaching post with that lawyer's "counterstory"). *See also* David Luban, *Difference Made Legal: The Court and Dr. King*, 87 MICH. L. REV. 2152, 2156 (1989) (contrasting two legal stories about the same set of demonstrations in order to show how "the self-same event entails radically different legal consequences when it appears in different narratives").

9. I refer to the central narrative as "Dreiser's narrative" only to distinguish it from the other stories in the novel. I do not mean to imply that the novel cannot be discussed without mention of Dreiser or that it is completely consistent with his expressed intent in writing it. For some of the ways in which *An American Tragedy* may undermine its dominant, Dreiser-sanctioned purpose, see *infra* text accompanying notes 36-39.

offered in the trial scene by the prosecution and the defense. Although the stories told at trial are ostensibly concerned only with Clyde's legal guilt or innocence, they seek to make the legal conclusions they justify morally palatable by affirming (in the case of the prosecution) or undermining (in the case of the defense) Clyde's general blameworthiness.[10] Like the defense's story, Dreiser's narrative encourages the reader to question whether Clyde can really be blamed for his behavior as a whole and for Roberta's death in particular. Its account of Roberta's death is, however, factually closer to that of the prosecution than to that of the defense. The split allegiances of Dreiser's narrative force the reader to recognize that the propriety of blame is not a simple factual judgment. The juxtaposition of the stories thus raises the question of how they manage to lead toward opposite conclusions regarding Clyde's blameworthiness.

As a general rule we feel that blame should not be assigned in the absence of responsibility.[11] We tend to find evidence of responsibility in the fact that someone acted, to use Hume's phrase, "according to the determinations of the will,"[12] or, colloquially, on purpose—a fact which we are inclined to equate with her will (and, less directly, the desire or decision that prompted it) having caused the action.[13] Even when someone has acted on purpose, however, we

---

10. Although we might hope that law and morality always coincide, we tend to distinguish the concept of legally sanctioned blame, which can depend on "considerations of policy and purpose," from the notion of morally justified blame, which requires assessment of whether the defendant was *really* to blame. *Cf.* JOEL FEINBERG, DOING AND DESERVING 30 (1970) (distinguishing between our concepts of legal and moral responsibility in these terms).

11. *See* JONATHAN GLOVER, RESPONSIBILITY 1 (1970).

12. DAVID HUME, ENQUIRIES CONCERNING THE HUMAN UNDERSTANDING AND CONCERNING THE PRINCIPLES OF MORALS 95 (1963). A vast number of the English and American philosophers who have discussed the subject agree that a determining will or voluntariness is a prerequisite for responsibility. *See* HONDERICH, *supra* note 7, at 451-87; *see also* Jonathan Bennett, *Accountability*, *in* PHILOSOPHICAL SUBJECTS 16 (Zak Van Straaten ed., 1980) (describing conditions for accountability).

13. Ted Honderich provides support for the general point that we ordinarily take there to be a causal relationship between mental events and actions, and for the more specific proposition that we take an action to be caused by the "active intention which represents it." HONDERICH, *supra* note 7, at 244 (emphasis omitted). He defines the concept of active intention as the execution or carrying out of a committed desire to do something, a belief about how to do it, and a belief that one can do it. *Id.* at 216-25. To say that we often treat a mental term as the cause of action, however, is not to claim that this treatment makes sense. The correctness of the causal picture of the relationship between mental events and actions is a matter of philosophical controversy. *Compare* Anthony Kenny, *Freedom, Spontaneity and Indifference*, *in* ESSAYS ON FREEDOM OF ACTION 87, 91 (Ted Honderich ed., 1973) (arguing that "nowadays . . . most philosophers would regard it as incorrect to think of wants as mental acts which determine action") *with* HONDERICH, *supra* note 7, at 248 (laying out his "Hypothesis on the Causation of Actions": "Each action is a sequence of bodily events which is the effect of a causal sequence

tend to question her responsibility if it seems she could not have helped choosing to act as she did.[14] Our notions of responsibility thus dictate that a story portraying human beings as purposive actors who are free to decide on the purposes they embrace will make its reader feel confident about ascribing responsibility and blame.[15] On the other hand, a narrative can encourage its reader to doubt the propriety of blame by making it seem as though all acts, including mental acts, are causally necessitated.[16] The first picture or model of human action is known, at least in legal circles, as intentionalism,[17] the second as determinism.[18]

I will argue that Dreiser's narrative, which, for the reader, describes what really happened, erodes Clyde's blameworthiness by presenting his actions and the world in which they occur in determinist terms. By contrast, the prosecution's intentionalist story encourages the reader to believe it is just to attach criminal blame to Clyde by depicting his behavior as entirely attributable to his free will. The defense's tale follows Dreiser's narrative in portraying Clyde as more acted upon than acting, but it alters the plot of the story so that

---

one of whose initial elements and some of whose subsequent elements are psychoneural pairs which incorporate the active intention which represents the sequence of bodily events.").

14. *See* HONDERICH, *supra* note 7, at 451-87; ANTHONY KENNY, FREEWILL AND RESPONSIBILITY 29, 34 (1978); KLEIN, *supra* note 7, at 1 ("[The] anxiety which comes naturally to us when we are asked to reflect on the conditions for moral responsibility . . . can be summed up in the question: how can someone be morally responsible for his acts if he is not responsible for the desires and beliefs which motivate him?"); Bennett, *supra* note 12, at 16; Roderick W. Chisholm, *Responsibility and Avoidability*, *in* DETERMINISM AND FREEDOM 157 (Sidney Hook ed., 1958).

15. Mark Kelman has noted that this sort of account of human behavior "renders the blaming practice morally unobjectionable." KELMAN, GUIDE, *supra* note 6, at 89.

16. The proposition that, to borrow Richard Taylor's formulation, "everything whatever is caused, and not one single thing could ever be other than exactly what it is," has long been felt to threaten our notions of blame and responsibility. RICHARD TAYLOR, METAPHYSICS 35 (1963) ("There is no moral blame nor merit in any man who cannot help what he does."). *See also* GLOVER, *supra* note 11, at 2 (discussing view that "moral responsibility is bound up with man not being a 'mere machine', so that it would be undermined if psychologists, neurophysiologists, and others could provide mechanistic models giving adequate causal explanations of all human behavior"). Philosophical discussions of whether this proposition is an actual or merely a perceived threat to our concepts of blame and responsibility have, as Hobbes noted over three hundred years ago, filled "vast and involuble volumes," HONDERICH, *supra* note 7, at 452, and, as I noted earlier, *supra* note 7, they continue to take place.

17. *See* KELMAN, GUIDE, *supra* note 6, at 86.

18. *See* Gary Watson, *Introduction*, *in* FREE WILL 2 (Gary Watson ed., 1982). For similar, broad definitions of determinism, see KELMAN, GUIDE, *supra* note 6, at 86; TAYLOR, *supra* note 16, at 34; PERRY O. WESTBROOK, FREE WILL AND DETERMINISM IN AMERICAN LITERATURE ix (1979); William James, *The Dilemma of Determinism*, *in* THE WILL TO BELIEVE AND OTHER ESSAYS IN POPULAR PHILOSOPHY 114, 117 (William James ed., 1979).

Clyde clearly did nothing for which the law would punish him.

The discussion that follows will begin by describing the way in which the structure and language of *An American Tragedy* convey a determinist picture of human behavior. I will then show how Dreiser's portrayal of his central character, Clyde Griffiths, contrasts with that offered by the prosecution. Finally, I will discuss the way in which the defense lawyers find themselves constrained in constructing Clyde's story by the intentionalist predispositions of the criminal law.

I

The plot of *An American Tragedy* is straightforward:[19] A young man with social and material ambitions brings about the death of his pregnant girlfriend because she stands in the way of his making a more socially desirable marriage. Dreiser was convinced that this type of murder happened "with surprising frequency" in America.[20] The Stuart case supports the argument that his conviction was justified.[21] Dreiser had become interested in similar crimes in 1892[22] and had attempted two novels and a short story, each based on a different murder, before publishing *An American Tragedy* in 1925.[23] It, too, is based on an actual murder: Chester Gillette's 1906 drowning of Grace "Billie" Brown at Big Moose Lake in the Adirondacks.[24] Although he altered the individuals and events of the Gillette case,[25] Dreiser borrowed heavily from both Grace Brown's letters, which

---

19. In his 1926 review of *An American Tragedy*, H.L. Mencken described the simple plot of the novel in pejorative terms as "[h]ardly more . . . than the plot of a three page story in *True Confessions*." H.L. Mencken, *Dreiser in 840 Pages*, *in* AMERICAN MERCURY, March 1926, at 379-81 *reprinted in* THOMAS P. RIGGIO, DREISER-MENCKEN LETTERS 797-98 (1986).

20. ROBERT H. ELIAS ET AL., LETTERS OF THEODORE DREISER 457-58 (1959) [hereinafter LETTERS] (letter to Jack Wilgus, April 20, 1927). The American public took Dreiser's insight seriously: Ten years after publication of *An American Tragedy*, when Robert Edwards committed this sort of crime, Dreiser was asked to be a special reporter at the trial. F.O. MATTHIESSEN, THEODORE DREISER 201 (1951).

21. When Charles Stuart, a former short-order cook from a poor background who had become a $100,000 per year fur salesman, shot his wife Carol in Boston, his reported motivation for the shooting was that Carol's pregnancy interfered with his hopes of an alliance with a woman of higher social standing. *See* Richard Lingeman, *Another American Tragedy*, N.Y. TIMES, Jan. 22, 1990, at A15 (noting the similarity between the Stuart case and *An American Tragedy*).

22. *See* Theodore Dreiser, *I Find the Real American Tragedy*, *reprinted in* Donald Pizer ed., Theodore Dreiser: A Selection of Uncollected Prose 291 (1977).

23. *See* DONALD PIZER, THE NOVELS OF THEODORE DREISER 205 (1976).

24. *See* ROBERT P. WARREN, HOMAGE TO THEODORE DREISER 98 (1971).

25. *See* RICHARD LEHAN, THEODORE DREISER: HIS WORLD AND HIS NOVELS 149 (1969); PIZER, *supra* note 23, at 217-22 (discussing specific alterations).

had been published in pamphlet form, and the extensive reports of Gillette's trial and execution published by the *New York World*.[26]

In the years following publication of *An American Tragedy*, Dreiser wrote a number of letters and articles in which he explained his fascination with the type of murder Gillette committed and his purpose in writing the novel.[27] Dreiser viewed this kind of crime as the manifestation of a distinctively American pattern of cause and effect.[28] American society, Dreiser wrote, "bred the fortune hunter de luxe" by valuing so highly the accumulation of wealth.[29] Chester Gillette's attempt to connect himself with the rich and socially privileged by marrying well, the success of which was threatened by Grace Brown's pregnancy, was really "*the kind of thing that Americans should and would have said was the wise and moral thing to do had he not committed a murder.*"[30] Gillette, as Dreiser saw him, was driven to murder by the conjunction of his "romantic dreams" and "those dreadful economic, social, moral, and conventional pressures about him":

> Not Chester Gillette, as I said to myself at the time, planned this crime, but circumstances over which he had no control—circumstances and laws and rules and conventions which to his immature and more or less futile mind were so terrible, so oppressive, that they were destructive to his reasoning powers.[31]

---

26. PIZER, *supra* note 23, at 215-16, 226-27.

27. *See, e.g.*, LETTERS, *supra* note 20, at 457-58 (letter to Jack Wilgus, April 20, 1927); LETTERS, *supra* note 20, at 509-10 (letter to Samuel Hoffenstein, February 26, 1931); LETTERS, *supra* note 20, at 510-12 (letter to Jesse L. Lasky, March 10, 1931); LETTERS, *supra* note 20, at 526-30 (letter to Harrison Smith, April 25, 1931). Dreiser wrote the last three of these letters during the course of his 1931 dispute with Paramount over the script for the film version of *An American Tragedy*. After Paramount failed to change the script as he wished, Dreiser sued. Although he lost the suit, Dreiser did get Paramount to add seven scenes to the film. *See* LETTERS, *supra* note 20, at 562.

28. In his letter to Jack Wilgus, Dreiser wrote:

> I had long brooded upon the story, for it seemed to me not only to include every phase of our national life—politics, society, religion, business, sex—but it was a story so common to every boy reared in the smaller towns in America. It seemed so truly a story of what life does to the individual—and how impotent the individual was against such forces.

LETTERS, *supra* note 20, at 457-58 (letter to Jack Wilgus, April 20, 1927).

29. DREISER, *supra* note 22, at 292. *See also* WARREN, *supra* note 24, at 137 (quoting Dreiser's *A Hoosier Holiday*, which describes the atmosphere of American cities as filled by a "'crude, sweet illusion about the importance of things material.'").

30. DREISER, *supra* note 22, at 297 (emphasis in original).

31. *Id.* at 299. Dreiser, as he often proclaimed in his essays, saw the universe, and not simply this crime, in determinist terms. *See, e.g.*, Theodore Dreiser, *The Essential Tragedy of Life*, *in* HEY-RUB-A-DUB-DUB 243 (Theodore Dreiser ed., 1920) (stating that man's "every move and aspiration [are] anticipated and accounted for

Dreiser intended *An American Tragedy* to set forth this determinist explanation of the Gillette genre of murders. The novel, he wrote, "is a progressive drama. . . . A certain and given chain of events leads to certain conclusions . . . ."[32] The reader must see Clyde as "a creature of circumstances" who is compelled to act as he does by "an inescapable web."[33]

I will argue that *An American Tragedy* successfully conveys the impression that the embedded values of our culture and our socio-economic, genetic, and psychological circumstances are the ultimate authors of our actions.[34] The novel does not, of course, address the difficult issue of when and how these different sorts of influences interact, and I do not wish to imply that its determinism is philosophically coherent. As Ted Honderich has observed: "It is one thing to declare or more likely to presuppose . . . that our choices and decisions are quite clearly effects, or that our behaviour is law-like. . . . It is another thing to set out a determinist theory which is explicit, complete, and at a proper level of specificity."[35] Nor do I wish to suggest that the novel describes every character and event in determinist terms. The degree to which some rich characters appear to be in charge of their lives[36] and the visible role of accidents in

---

by a formula"). For discussions of the sources on which Dreiser drew in formulating his ideas see RONALD E. MARTIN, AMERICAN LITERATURE AND THE UNIVERSE OF FORCE 217-36 (1981); PIZER, *supra* note 23, at 211-13. Although Dreiser "persevered in his philosophizing as though he were on the brink of some great discovery," MARTIN, *supra* this note, at 218, he was criticized not only for the incoherence of his views but also for their seeming inconsistency with the energy he poured into social causes. *See, e.g.*, LETTERS, *supra* note 20, at 784-85 (letter to Robert H. Elias, April 17, 1937, noting this criticism and attempting to reconcile his determinism and his desire to take part in social reforms).

32. *See also* LETTERS, *supra* note 20, at 510-12 (letter to Jesse L. Lasky, March 10, 1931).

33. *Id. See also* LETTERS, *supra* note 20, at 526-30 (summarizing Dreiser's "ideographic plan" for the novel).

34. That *An American Tragedy* is a determinist novel is something of a literary truism. *See* SUSAN L. MIZRUCHI, THE POWER OF HISTORICAL KNOWLEDGE 242-43 (1988). Some of the more interesting recent analyses of the novel's determinism appear in LEE C. MITCHELL, DETERMINED FICTIONS 55-74 (1989) (focusing on the narrative techniques by which Dreiser renders determinism real); MIZRUCHI, *supra* this note, at 242-94 (discussing the ways in which the novel reflects critically on the social uses of and undermines determinism); MICHAEL SPINDLER, AMERICAN LITERATURE AND SOCIAL CHANGE (1983) (extracting a strictly economic determinism from *An American Tragedy* and arguing that Clyde is trapped by the transition from a consumption-oriented to a production-oriented society).

35. HONDERICH, *supra* note 7, at 3.

36. Susan Mizruchi has argued that the novel affords the wealthy a capacity to "manipulate temporal perceptions and historical narratives" which challenges its dominant philosophy of determinism. MIZRUCHI, *supra* note 34, at 269. Mark Kelman has remarked on a similar tendency to identify privilege with responsibility and blameworthiness in the criminal law. *See* KELMAN, GUIDE, *supra* note 6, at 90.

shaping the course of Clyde's life[37] can both undermine the idea that everything is predetermined. And at least one critic has read the section of the novel following Clyde's trial—in which Clyde, with Reverend MacMillan's help, concludes that he was not "wholly white"[38] regarding Roberta's death—as revealing that Dreiser afforded Clyde a measure of personal guilt.[39] Yet, because the novel creates such an overwhelming sense of "causal sequence and inevitability,"[40] these (arguably) inconsistent moments do not prevent it from privileging the general principle that human behavior is the causal consequence of conditions the individual does not influence.

Although *An American Tragedy* contains the odd statement explaining a character's behavior as the product of a "chemism"[41] or a "fillip in the blood,"[42] it is remarkably free of the scientific and philosophical proclamations that characterize many of Dreiser's other

---

37. Some examples of decisive accidental events in Clyde's life are: the automobile accident that forces Clyde to run away from Kansas City; the chance meeting with his uncle that leads him to Lycurgus; and the unintentional blow he gives Roberta with his camera, pushing her into the water where she drowns. As Philip Fisher has noted, these accidents underscore the point that Clyde is not in control of his life. *See* Philip Fisher, *Looking Around to See Who I Am: Dreiser's Territory of the Self*, 44 ELH 728, 741 (1977); *see also* WESTBROOK, *supra* note 18, at 149. For a concise discussion of the way in which chance, like determinism, seems to undermine our control of and moral responsibility for our actions see Thomas Nagel, *Moral Luck*, *in* FREE WILL, *supra* note 18, at 174-86.

Although Dreiser's use of accident helps to remove Clyde from the center of his actions, it may, depending on the way in which it is interpreted, undermine *An American Tragedy*'s governing determinism. Insofar as the "accidental" events in the novel are accidental from the point of view of an omniscient narrator and thus symptomatic of a random universe, they will conflict with determinism's insistence on predetermined causal chains. *See* JAMES, *supra* note 18, at 117-18. Insofar as the events in question are portrayed as "accidental" from a particular rather than an omniscient point of view—say, from the author's or the central character's perspective—they need not be indicative of a lack of predetermination. Critics who wish to make Dreiser a consistent determinist tend to interpret the "accidents" in Clyde's story as only appearing to be accidental. *See, e.g.*, JOHN J. CONDER, NATURALISM IN AMERICAN FICTION 86-87 (1984) (quoting Dreiser's statement that the concept of chance is "'in most cases only another name for our ignorance of causes'" and arguing that the chance events in Dreiser's novels are not intended to challenge determinism); LEHAN, *supra* note 25, at 164-65 (claiming that the "accidents" that happen to Clyde are only accidental from Clyde's point of view).

38. THEODORE DREISER, AN AMERICAN TRAGEDY 807 (1981).

39. *See* LAWRENCE E. HUSSMAN, JR., DREISER AND HIS FICTION 133-35 (1983). Even if this last section of the novel does afford Clyde some measure of responsibility, my sense is that that measure is not sufficient to sustain blame.

40. LEHAN, *supra* note 25, at 164.

41. DREISER, *supra* note 38, at 13. Dreiser began to use the term "chemism" after studying the ideas of Freud and Loeb during the second decade of the twentieth century. *See* PIZER, *supra* note 23, at 211-12.

42. DREISER, *supra* note 38, at 196.

works.[43] Thus, it has in common with other stories whose authors are less consciously loyal to a particular set of ideas that it makes the views that inform it plausible through narrative technique rather than through direct assertion. *An American Tragedy* privileges determinism through concrete description, prose style, and narrative strategies such as repetition and foreshadowing. I will first summarize Dreiser's narrative in an effort to show that his descriptions lead the reader to look to the messages of American society and to Clyde's socioeconomic, genetic, and psychological circumstances, rather than to his will, for an explanation of his behavior. I will then discuss the various ways in which Dreiser reinforces the insignificance of Clyde's will in producing his actions. Finally, I will describe Dreiser's use of particular prose constructions, repetition, foreshadowing, and impersonal descriptions to make the reader feel the remorseless operation of causal laws.

The most immediately striking feature of *An American Tragedy* is its length. The novel, which contained one million words in manuscript,[44] finally weighed in at three hundred and eighty-five thousand.[45] There are, of course, many factors contributing to its "gargantuan"[46] size. At least two of these factors—the broad time frame[47] in which Dreiser chooses to set the story of the murder and the detail with which he describes scenes, characters, and events[48]—are important in making causal analysis of Clyde's thoughts and actions possible.

An American Tragedy is divided into three books, the first of which concerns Clyde's early adolescence in Kansas City. The novel opens from the perspective of a largely indifferent pedestrian strolling on a summer's evening through the commercial center of a large American city. The reader observes the Griffiths family as a visibly impoverished, unnamed family of six who preach on a street corner. She notes that the eldest son of the family looks restless and uncomfortable with his missionary role, "eyes down, and for the most part only half singing."[49] The reader's first response to the boy, whom she will shortly know as Clyde, is articulated by one of the strangers

---

43. *See* MARTIN, *supra* note 31, at 253.

44. *See* DREISER-MENKEN LETTERS, *supra* note 19, at 506.

45. *See* MENCKEN, *supra* note 19, at 797.

46. *Id.* at 799.

47. Mark Kelman has observed that choosing a broad time frame in which to tell the story of the defendant's actions provides enough "background data," enough facts about events preceding the criminal incident, to make determinist explanation possible. *See* Kelman, *Interpretive Construction*, *supra* note 6, at 594.

48. Robert Penn Warren has noted that there is a "dawning logic" in the novel's "scrupulous accretion of detail, small indications, and trivial events." Robert P. Warren, *An American Tragedy*, LII No. 1 THE YALE REVIEW 6 (1962).

49. DREISER, *supra* note 38, at 9.

who observes him: "'That oldest boy don't wanta be here. He feels outa place, I can see that. It ain't right to make a kid like that come out unless he wants to.'"[50]

From the beginning, then, the novel asks the reader to observe the effect on Clyde of circumstances forced upon him. Dreiser makes it clear that Clyde is deeply affected by the "wretched and hum drum, hand to mouth state" in which he spends his childhood.[51] Clyde's itinerant missionary parents do not provide their children with sufficient food, decent clothes, or even the chance to go to school on a regular basis. The only real lesson Clyde learns from his childhood is that his parents' work is "not satisfactory to others—shabby, trivial" and that he must dissociate himself from them if he wishes not to be treated with contempt.[52]

The novel shows the reader that Clyde's appearance and temperament join with the shame and poverty of his childhood to produce a burning ambition to get away and to better himself. The adjectives Dreiser uses to describe Clyde convey the impression that he is vain, proud, imaginative, self-absorbed, highly-sexed, and indecisive.[53] Clyde is good-looking, with "a straight, well-cut nose, high white forehead, wavy, glossy black hair, [and] eyes that [are] black and rather melancholy at times."[54] Because he is aware of his good looks, he is resentful of the financial circumstances that keep him from buying the "right" clothes.

By his sixteenth birthday Clyde is able to begin the "race" to succeed in life. Eager to begin, he quits school and accepts a job as assistant to a soda water clerk at one of the cheaper drugstores in Kansas City. He progresses from there to the post of bell-hop at the "principal hotel of the city,"[55] the Green-Davidson. Dreiser leads the reader to believe the exposure to this "gaudy"[56] hotel, which lacks "the saving grace of either simplicity or necessity"[57] injures Clyde's development by providing him with an exclusively material picture of success. He also shows the reader that the effect of the Green-Davidson on Clyde is not something for which Clyde himself is responsible. Clyde shares with the other bell-boys the fact that he comes from a nondescript family without social or financial advantages. The common background of the bell-boys produces in them a common response to the Green-Davidson:

---

50. *Id.* at 11.
51. *Id.* at 17.
52. *Id.* at 14.
53. DREISER, *supra* note 38, at 18.
54. *Id.*
55. *Id.* at 31.
56. *Id.* at 47.
57. DREISER, *supra* note 38, at 32.

> [T]here had been in the lives of most of these boys such a lack of anything that approached comfort or taste, let alone luxury, that, not unlike Clyde, they were inclined to not only exaggerate the import of all that they saw, but to see in this sudden transition an opportunity to partake of it all.[58]

Unfortunately, Clyde's dreams of what his connections with the Green-Davidson might mean are quashed after a year of work. One of the bell-boys borrows a car from a friend who lends it without the owner's consent. Clyde joins a group of the bell-boys in taking the car for an outing in the country. On their return to Kansas City the speeding driver of the car hits and kills a little girl who has jumped into the street. Attempting to elude the policemen who are in pursuit, the driver crashes the car into a lumber pile. The first book of the novel ends with Clyde running from the car and feeling keenly aware of his dependence on forces he cannot control: "If he were not captured, he hoped to hide—to lose himself and so escape—if the fates were only kind . . . ."[59]

The second book of *An American Tragedy* opens with a description of dinner preparations at the home of Clyde's uncle, Samuel Griffiths, in Lycurgus, New York. Griffiths, the wealthy owner of a shirt collar factory, has met Clyde at the Union League Club in Chicago, where Clyde, after three years of wandering from city to city and job to job, has obtained a post as a bell-hop. Impressed by the "gentlemanly and reserved"[60] manner Clyde has acquired at the Union League Club, Samuel Griffiths grants Clyde's request for a job in the collar factory. Clyde, taken with the idea of connecting himself with the rich uncle he heard about as a child, quits his job at the Union League Club and goes to Lycurgus.

Samuel Griffiths, motivated by Clyde's resemblance to his own son, Gilbert, and by a sense of remorse at the straightened circumstances of Clyde's father, decides to teach Clyde "the collar business" "from top to bottom."[61] Clyde's first job at the factory is in the basement shrinking room. Unfortunately, Samuel Griffiths does not tell Clyde about his ambitions for him, and Clyde is left to guess about his status in the factory. He feels alone in Lycurgus, a "nobody"[62] who is isolated from the lofty Lycurgus Griffiths by his parents and his past and from the other factory workers by his ambitions and his Griffiths name.

After some months Samuel Griffiths puts Clyde in charge of a room full of young women who stamp size numbers on collars.

---

58. *Id.* at 48.
59. *Id.* at 145.
60. *Id.* at 169.
61. DREISER, *supra* note 38, at 175.
62. *Id.* at 189.

Despite company regulations forbidding employee relationships, Clyde becomes intimately involved with one of the employees he supervises, Roberta Alden. Like Clyde, Roberta is a stranger in Lycurgus and is from a poor background. Clyde, with his good looks and Griffiths name, embodies for Roberta the possibility of her own escape from the rural poverty of her childhood. She is attracted to him, just as Clyde is attracted to the Lycurgus Griffiths. After their relationship has continued for some months, Roberta yields to Clyde's sexual demands.

By the time Clyde learns that Roberta is pregnant, he has fallen in love with Sondra Finchley, heir to the Finchley Vacuum Cleaner fortune. Dreiser explains that the Finchleys, like the Green-Davidson, display their wealth in a "grandiose" and "showy" manner.[63] Although at first Sondra is interested in Clyde only to make his cousin Gilbert jealous, she becomes genuinely attached to him and includes him in her social activities. As Sondra becomes increasingly interested in Clyde, giving Clyde reason to hope she will marry him, Clyde loses all interest in Roberta. His efforts to help Roberta abort the baby fail, and the thought of marriage, which Roberta now demands, fills him with the fear of losing his job and "all the joys that so recently in connection with Sondra had come to him."[64]

Prevented by his "mental and material weakness" from denying Roberta's request to marry her *and* from giving up the pleasures and dreams associated with Sondra, Clyde feels completely trapped.[65] At this point he happens to read about the "mysterious drowning" of a young couple at an isolated lake in Massachusetts. The couple's rented boat was found overturned with two hats floating nearby. Though the woman's body was discovered, the man's was not. In a conversation between Clyde and the "genie of his darkest and weakest side,"[66] the genie strives to convince Clyde to repeat the circumstances of the news story with himself and Roberta (who cannot swim) by drowning her and making it appear as though her anonymous male companion has perished as well.[67]

---

63. *Id.* at 149.

64. *Id.* at 436.

65. *See* DREISER, *supra* note 38, at 466.

66. *Id.* at 464.

67. The genie supplies Clyde with the following summary of his reasons for killing Roberta:

> "But a little blow—any little blow under such circumstances would be sufficient to confuse and complete her undoing. Sad, yes, but she has an opportunity to go her own way, has she not? And she will not, nor let you go yours. Well, then, is this so terribly unfair? And do not forget that afterwards there is Sondra—the beautiful—a home with her in Lycurgus—wealth, a high position such as elsewhere you may never obtain again—never—never. Love and happiness—the equal of any one here—superior even to your cousin Gilbert."

Id.

When Roberta announces that if Clyde does not marry her immediately she will reveal all to anyone in Lycurgus who cares to listen, Clyde is "finally all but numbed by the fact that now decidedly he must act."[68] He makes arrangements for Roberta to go to Grass Lake with him for what he tells her is a honeymoon trip after which they will be married. When Grass Lake turns out to be well-populated by the Winebrennarian Religious Group from Pennsylvania, Clyde persuades Roberta to accompany him to a more remote lake, Big Bittern. There, Clyde hires a boat and rows out onto the lake with Roberta; but once they are on the water, he is paralyzed by doubt and confusion. In a kind of "trance or spasm"[69] he experiences "a sudden palsy of the will"[70] and recognizes that he lacks the courage to drown Roberta as he had planned. Roberta, who is alarmed at the strange expression on Clyde's face, crawls forward toward him on the keel of the boat and reaches out to him. As she draws near, he flings out at her instinctively, wishing simply to prevent her touch. Clyde has forgotten that he is holding a camera in the hand he uses to reject Roberta. When he pushes her away, the camera strikes her; she screams and is thrown backward. Clyde rises and reaches across to her, "half to assist or recapture her and half to apologize for the unintended blow."[71] In rising he capsizes the boat. Realizing that an accident has accomplished what he sought yet lacked the will to do, he allows Roberta to drown. Book Two ends with Clyde swimming to shore, wondering whether Roberta's death was an accident or whether he has killed her.[72]

The first two books of *An American Tragedy* lead the reader to believe that Clyde *is* responsible for Roberta's death in the same direct, morally-neutral sense that the rain is responsible for the wet roof. The reader feels that Roberta would have lived were it not for Clyde's attempt to dispose of her. Yet, because Dreiser has consistently described Clyde as responding rather than originating, the reader is inclined to feel that Clyde is not the true author of Roberta's death. Dreiser's narrative teaches that Clyde becomes a murderer not

---

68. *Id.* at 469.
69. *Id.* at 492.
70. *Id.* at 491.
71. *Id.* at 492.
72. *Id.* at 494.

> And then Clyde, with the sound of Roberta's cries still in his ears, that last frantic, white, appealing look in her eyes, swimming heavily, gloomily and darkly to shore. And the thought that, after all, he had not really killed her. No, no, Thank God for that. He had not. And yet (stepping up on the near-by bank and shaking the water from his clothes) had he? Or, had he not? For had he not refused to go to her rescue, and when he might have saved her, and when the fault for casting her in the water, however accidentally, was so truly his? And yet—and yet—.

Id.

by choice but by the effect of the embedded values of American society on someone of Clyde's background and temperament.[73] Institutions like the representative Green-Davidson,[74] as interpreted by American culture and perceived by the deprived and unsophisticated Clyde, encourage him to equate personal worth with financial success[75] and to believe that wealth is within his reach.[76] Lacking the skills and the trained or naturally clear mind that would have allowed him to make money himself, Clyde perceives association with the affluent as his only means of achieving success. He relies successively on the Green-Davidson, the Union League Club, and the Lycurgus Griffiths to give him a positive sense of himself. Marriage to Sondra represents the possibility of cementing his connection to luxury and position, of sustaining a feeling of personal worth over time. Having observed the isolating and downtrodden circumstances of Clyde's youth, his vanity and pride, and the imaginative and confused character of his mind, the reader understands what this possibility means to him and perceives as inevitable his attempt to eliminate Roberta, whom he regards as the only obstacle to its realization.

Dreiser reinforces the impression that Clyde is not in charge of his life by making references throughout the novel to the fact that Clyde lacks a clear, directing will.[77] When he first considers murdering Roberta and so "solving" his dilemma, Clyde is unable to form a clear resolve to do so because of his "unstable and highly variable will."[78] "He could not really act on such a matter for himself and would not. It remained as usual for him to be forced either to act or to abandon this most *wild* and terrible thought."[79] Clyde does not choose independently to drown Roberta. He is persuaded to endorse the plan by his genie, a voice outside him.[80] And it is the

---

73. The novel's title, of course, reinforces the point that Clyde's fate is distinctively American.

74. Dreiser makes the Green-Davidson's archetypal status clear, referring to it as "an essential hotel in a great and successful American commercial city . . . ." *Id.* at 32.

75. DREISER, *supra* note 38, at 47.

76. *Id.* at 48.

77. *See, e.g.*, *id.* at 18, 169.

78. *Id.* at 467.

79. *Id.* (emphasis in original).

80. By using the genie to explain Clyde's attraction to the idea of drowning Roberta, Dreiser leaves intact our sense that Clyde is incapable of clear thought or resolution. Describing Clyde's state of mind when he listens to the genie, Dreiser writes:

> There are moments when in connection with the sensitively imaginative or morbidly anachronistic—the mentality assailed and the same not of any great strength and the problem confronting it of sufficient force and complexity—the reason not actually toppling from its throne, still totters or is warped or shaken—the mind befuddled to the extent that for the time

"eerie unreason or physical and mental indetermination" on Clyde's face that causes Roberta to start on her ill-fated crawl across the boat.[81]

Dreiser describes many of Clyde's reactions in a way that suggests diminished capacity and makes it difficult to perceive Clyde as in charge of his life. When Clyde gets his first paycheck at the Green-Davidson, his salary seems "fantastic, Alladinish, really."[82] Similarly, when he visits a brothel with the other Green-Davidson bell-boys, it is "really quite an amazing and Aladdin-like scene to him."[83] These phrases encourage the reader to believe that the presence of money or sex casts a spell over Clyde and that he is bewitched by Sondra Finchley, in whose person they are united.

The prose style of Dreiser's narrative further attests to the appropriateness of picturing Clyde's actions as the result of his ideological, socioeconomic, genetic, and psychological circumstances—rather than as the product of his will—by conveying the general impression that these circumstances determine behavior. Dreiser's tendency to omit the verbs from many sentences helps to portray characters as more caused than causing.[84] And his habit of describing characters as the loci of social and genetic forces, rather than as distinct individuals, encourages the reader to afford these forces the primary causal role. Clyde's father, for example, is described as "one of those poorly integrated and correlated organisms, the product of an environment and a religious theory but with no guiding mental insight of his own."[85]

As the paragraph above suggests, Dreiser's prose style helps to extend the reach of the novel's determinism beyond Clyde and others similarly situated.[86] Dreiser's style leads the reader to participate in Clyde's experience of powerlessness. His extensive use of the present participle, which makes actions seem less self-contained than do other

---

> being, at least, unreason or disorder and mistaken or erroneous counsel would appear to hold against all else. In such instances the will and the courage confronted by some great difficulty which it can neither master nor endure, appears in some to recede in precipitate flight, leaving only panic and temporary unreason in its wake.

DREISER, *supra* note 38, at 463.

81. *Id.* at 492.

82. *Id.* at 53.

83. *Id.* at 66.

84. *See, e.g.*, DREISER, *supra* note 38, at 7.

85. *Id.* at 13. For other examples of Dreiser's tendency to depict his characters as produced rather than self-fashioned see *id.* at 196, 504, 513. For a detailed discussion of the ways in which Dreiser erodes the importance of individual identity, see Fisher, *supra* note 37, at 740-41.

86. Lee Clark Mitchell has also explored the notion that Dreiser's "bad writing" is instrumental in conveying determinism. *See* MITCHELL, *supra* note 34, at x, 73-74.

verb forms, and of long sentences strung together with dashes and commas, traps the reader within the text and gives her the sense of being pushed along helplessly. Dreiser's style also helps to create an atmosphere of inevitability. His practice of juxtaposing sentences of parallel construction[87] conveys a sense that the feelings, actions, or events the sentences describe occur according to fixed patterns.

Dreiser repeats not only sentence constructions but types of characters and events.[88] As Lee Clark Mitchell has observed, Clyde's and Roberta's fathers are "virtual carbon copies" of each other, befuddled and incompetent younger sons.[89] Clyde and Roberta represent each other other in important ways. Both are romantically inclined young people from isolating and deprived backgrounds and are attracted to the idea of marrying someone more privileged.[90] And many of the events in Lycurgus reflect those in Kansas City.[91] The event of unwanted pregnancy, for example, traps Clyde's sister Esta in Book One and then Roberta in Book Two.

Dreiser uses anticipation as well as repetition to reinforce the impression that our lives proceed according to a predetermined logic.[92] While Clyde is working at the Union League Club, he notices that this institution, which, unlike the Green-Davidson, caters to the truly "socially worldly elect," is completely without "the faintest trace of that sex element which had characterized most of the phases of life seen at the Green-Davidson."[93] Clyde's recognition of the absence of sex, and women, from the Club leads him to reflect: "Probably one could not attain to or retain one's place in so remarkable a world as this unless one were indifferent to sex, a disgraceful passion of course."[94] Dreiser then informs the reader that when Clyde "was within the precincts of the club itself, he felt himself different from what he really was—more subdued, less romantic, more practical . . . ."[95] Clyde's perception that lasting membership in the elite class requires an ascetic temperament and Dreiser's comment that Clyde lacks such a temperament together anticipate that Clyde's sexual desires, which ultimately lead him into Roberta's bed, will

---

87. Donald Pizer has discussed this practice, which he interprets primarily as a means of communicating the stream of consciousness or the "swift, uncontrolled flow of interior reality." PIZER, *supra* note 23, at 287-88.

88. Fuller discussions of Dreiser's use of repetition appear in LEHAN, *supra* note 25, at 164-67; MITCHELL, *supra* note 34, at 57-70; PIZER, *supra* note 23, at 282-83.

89. MITCHELL, *supra* note 34, at 59.

90. *See id.*

91. *See* LEHAN, *supra* note 25, at 164.

92. For critical discussions of Dreiser's use of foreshadowing see LEHAN, *supra* note 25, at 165-66; MITCHELL, *supra* note 34, at 65.

93. DREISER, *supra* note 38, at 168.

94. *Id.* at 169.

95. *Id.*

prevent him from realizing his dream of success. Here, as elsewhere, one scene foreshadows another, implying that determinism is true, that "those parts of the universe already laid down absolutely appoint and decree what the other parts shall be."[96]

A determined universe is, of course, a universe whose character results from the operation of impersonal causal laws rather than from the sum of individual actions. A sense of impersonality pervades Dreiser's narrative. Its first sentence refers not to the particular but to the generic: "Dusk—of a summer night. And the tall walls of a commercial heart of an American city of perhaps 400,000 inhabitants. . . ."[97] And characters are consistently described not as individuals but as types. The actor with whom Clyde's sister runs away is "one of those vain, handsome, animal personalities, all clothes and airs but no morals."[98] Roberta's parents are "excellent examples of that native type of Americanism which resists facts and reveres illusion."[99] Roberta's roommate, Grace Marr, is "of that type that here as elsewhere find the bulk of their social satisfaction in such small matters as relate to the organization of a small home . . . ."[100] Dreiser thus conveys the impression that a character's individuality is not a necessary element of her role in the novel's events.[101]

In these ways Dreiser creates a context of "grinding impersonality"[102] for Clyde's story. This context or atmosphere lends plausibility to Dreiser's rhetorical attribution of Clyde's crime to the pressures of American convention and to Clyde's psychological and genetic make-up and deprived background. It also leads the reader to believe that the predetermined nature of Clyde's conduct is the norm rather than an exception.

II

In contrast to Dreiser's narrative, which instills the feeling that Clyde is not ultimately responsible for Roberta's death, District

---

96. JAMES, *supra* note 18, at 117.

97. DREISER, *supra* note 38, at 7.

98. DREISER, *supra* note 38, at 21.

99. *Id.* at 244.

100. *Id.* at 247-48.

101. The large number of characters in *An American Tragedy*, the fact that each book of the novel begins with a description in general terms by a detached observer, and the shifts in perspective asking the reader to contrast Clyde's importance in his own mind with the indifference he inspires in others all serve to reinforce the reader's sense of the insignificance of individuals. For illustrations of the way in which Dreiser shifts his distance from Clyde at different points in the novel, thus showing Clyde "in his double aspect" as a "confused sufferer and victim of fate," see Howe, *Afterward*, *in* DREISER, *supra* note 38, at 825.

102. WARREN, *supra* note 24, at 113. Even a reviewer as negative as H.L. Mencken attested to *An American Tragedy*'s atmosphere of inevitability: "What we behold is the gradual, terrible, irresistable approach of doom—the slow slipping away of hopes. The thing somehow has the effect of a tolling of bells. It is clumsy. It lacks all grace. But it is tremendously moving." Mencken, *supra* note 19, at 799.

Attorney Mason's version of Clyde's story makes it seem as though Roberta's death is the direct result of Clyde's malicious and deliberate actions. Mason, who does not have the reader's detailed knowledge of the circumstances of Roberta's death, must reconstruct them from the available evidence. Roberta's body, with bruises on the face, is dragged from the lake at Big Bittern.[103] Her bag is found at the lodge, but that of the companion with whom the guide and innkeeper saw her is missing.[104] The couple is found to have registered in one name at Grass Lake and in another at Big Bittern, and the guide who drove them to Big Bittern remembers Roberta's companion as having asked whether there were many people on the lake that day.[105] A letter addressed to her mother is found in Roberta's coat.[106] Mason contacts her parents, who suggest that Clyde Griffiths of Lycurgus could have been her companion.[107] Mason obtains access to Clyde's room in Lycurgus and discovers three bundles of letters: one from Roberta, one from Sondra, and one from Clyde's mother.[108] Roberta's letters, written from her parents' house where she had gone to "rest" for a few weeks before joining Clyde at Grass Lake, reveal her pregnancy and anxious desire to marry Clyde.[109] Sondra's opulent notecards and invitations contrast mightily with Roberta's letters in both tone and appearance.[110] On the basis of this evidence, Mason arrests Clyde.[111] Shortly after Clyde's arrest, divers combing the bottom of Big Bittern discover his camera.[112]

As the reader watches Mason attempt to derive from this evidence the "true" story of Roberta's death, she gains insight into the extent to which factors independent of the evidence shape legal narrative. Dreiser reveals that Mason's background and psychology predispose him to resent Clyde. Before reading about Mason's investigation of Roberta's death, the reader learns he is the son of a poor farmer's widow whose boyhood of "poverty and neglect" has caused him "to look on those with whom life has dealt more kindly as too favorably treated."[113] Although he has managed to escape the straightened circumstances of his youth, Mason remains a man of "limited social experience."[114] Mason's lack of sophistication leads him to impose oversimple interpretations on the character and behavior of those he

---

103. DREISER, *supra* note 38, at 498.
104. *Id.* at 499.
105. *Id.*
106. *Id.* at 499-500. The letter reveals her intent to marry Clyde. *Id.*
107. *See* DREISER, *supra* note 38, at 509-18.
108. *See id.* at 522-26.
109. *Id.* at 524.
110. *Id.* at 525. Sondra's letters are perfumed and on monogrammed stationary, whereas Roberta's are "doleful" and "pathetic." *Id.* at 525-26.
111. *Id.* at 526, 553.
112. DREISER, *supra* note 38, at 575.
113. *Id.* at 504.
114. *Id.* at 525.

investigates.[115] Moreover, Mason is inclined to resent those who appear to have been successful with women. He broke his nose at the age of fourteen, and the resulting disfigurement "had eventually resulted in what the Freudians are accustomed to describe as a psychic sex scar."[116]

If Mason's history and temperament thus lead him to resent Clyde, who bears the wealthy Griffiths name and has not one but two women pursuing him, it is the hierarchy of crimes embedded in American law and culture that creates in him a need to find that Clyde has willfully murdered Roberta according to a premeditated plan. The most severe legal and moral sanctions attach to such a killer. Mason, who is prone to think the worst of Clyde, is thus likely to view his behavior in intentionalist terms. Mason's legal ambitions harden this predisposition into a desire to find Clyde guilty of first degree murder. Securing a well-publicized conviction for such a heinous crime could, Mason knows, improve his chances of success in the local judicial elections.[117]

The evidence Mason uncovers is consistent with the proposition that Clyde planned and willfully caused Roberta's death. The letters from Sondra and Roberta supply motive; Clyde's suspicious behavior at Grass Lake and Big Bittern suggests premeditation; and the correspondence between the measurements of Clyde's camera and those of the bruises on Roberta's face indicates he hit her with the camera before she drowned. Clyde, unaware that divers discovered his camera, insists to Mason that he did not strike Roberta and does not in any way suggest that he did strike her but unintentionally. Mason, therefore, is free to conclude that Clyde deliberately beat Roberta unconscious with the camera and then threw her in the lake to drown.

In his opening statement to the court Mason states his intention to prove that Clyde killed Roberta "with malice aforethought and in cold blood . . . ."[118] He then proceeds to sketch Clyde's background and character so that they appear consistent with premeditated, willful killing. It is worth quoting from Mason's narrative at length in order to give a sense of its contrast with Dreiser's:

> But who is the individual . . . against whom I charge all these things? . . . Is he the son of wastrel parents—a product of the slums—one who had been denied every opportunity for a proper or honorable conception of the values and duties of a decent and respectable life? Is he? On the contrary. His father is the same strain that has given Lycurgus one of its largest and most construc-

115. *Id.*
116. *Id.* at 504.
117. *See* DREISER, *supra* note 38, at 508.
118. *Id.* at 640.

> tive industries . . . [h]is parents . . . appear to have been unordained ministers of the proselytizing and mission-conducting type—people who, from all I can gather, are really, sincerely religious and right-principled in every sense. . . .
>
> He is not a boy. He is a bearded man. He has more social and educational advantages than any one of you in the jury box. He has traveled. In hotels and clubs and the society with which he was so intimately connected in Lycurgus, he has been in contact with decent, respectable, and even able and distinguished people. Why, as a matter of fact, at the time of his arrest two months ago, he was part of as smart a society and summer resort group as this region boasts. Remember that! His mind is a mature, not an immature one. It is fully developed and balanced perfectly.[119]

Mason does not lie about Clyde's past. He does, however, omit all facts that might lead the jury to sympathize with Clyde and to see his youth as downtrodden and desperate, and he emphasizes those that link him with privilege. Mason relies on this selection of facts to substantiate his claim that Clyde's mind is "mature" and "fully developed," the sort of clear, rational mind possessed by someone who knows how to get what he wants. To follow Mason's reasoning, then, the jury need endorse only a selective intentionalism, according to which those with privilege have the capacity to control their actions. Even a selective intentionalism conflicts, however, with the thorough-going mechanism conveyed by Dreiser's narrative.[120]

III

Clyde's lawyers, Alvin Belknap and Reuben Jephson, offer a portrait of Clyde's background and character which evokes a selective but unexceptional determinism that may be said to be consistent with Mason's selective intentionalism. Its contention is that Clyde is the sort of person who cannot control his actions or desires. In direct contrast to Mason's assertion that Clyde is a "bearded man" of "social and educational advantages,"[121] Belknap and Jephson emphasize Clyde's youth and the extent to which he has been affected by the deprivations of his childhood. Like Dreiser, Belknap, in his

---

119. *Id.* at 641-42.

120. It would be possible to argue that, although Dreiser and Mason clash over whether Clyde is responsible for his conduct, their underlying theories of human action do not conflict. As I noted earlier, Dreiser is open to the charge that the behavior of poor people in *An American Tragedy* is more explicitly portrayed as determined than that of the wealthy. *See supra* note 36. If, as I have argued, the dominant impression Dreiser's narrative conveys is one of a mechanical universe, then the nature of human conduct, and not simply the nature of Clyde's background and behavior, is at issue in the competition between Dreiser's story and Mason's narrative.

121. *See* DREISER, *supra* note 38, at 641-42.

closing statement, suggests that Clyde did not possess the power to alter his behavior toward Roberta:

> [I]t became Belknap's duty to say his last word for Clyde. And to this he gave an entire day, most carefully, and in the spirit of his opening address, retracing and emphasizing every point which tended to show how, almost unconsciously, if not quite innocently, Clyde had fallen into the relationship with Roberta which had ended so disastrously for both. Mental and moral cowardice, as he now reiterated, inflamed or at least operated on by various lacks in Clyde's early life, plus new opportunities such as previously had never appeared to be within his grasp, had affected his "*perhaps too pliable and sensual and impractical and dreamy mind.*"[122]

Belknap's reference to Clyde's "mental and moral cowardice" is reminiscent of Dreiser's description of Clyde's "mental and material weakness before pleasures and dreams he could not bring himself to forego."[123] Similarly, Jephson's cross-examination of Clyde reminds the reader of Dreiser's references to Clyde's bewitchment by the "Alladinish" attributes of money and sex:

> "Yes, this Miss X.[124] We know. You fell madly and unreasonably in love with her. Was that the way of it?" "Yes, Sir." "And then?" "Well . . . I just couldn't care for Miss Alden so much any more." A thin film of moisture covered Clyde's forehead and cheeks as he spoke. "I see! I see!" went on Jephson, oratorically and loudly. . . . "A case of the Arabian Nights, of the enscorcelled and the enscorceller." "I don't think I know what you mean," said Clyde. "A case of being bewitched, my poor boy—by beauty, love, wealth, by things that we sometimes think we want very, very much, and cannot ever have . . . ."[125]

Belknap and Jephson thus attempt to convince the jurors that it would not be just to blame Clyde for Roberta's death by asking them to recognize that when "we" are young and deprived we often have no power to resist the lure of beauty, love, or wealth. It would be the exceptional disadvantaged youth who had the capacity to resist such things.

Although Clyde's lawyers rely on the rhetoric and selection techniques of determinist discourse to suggest that Clyde is morally

---

122. *Id.* at 734 (emphasis in original).

123. *Id.* at 466-67.

124. The lawyers agree not to reveal Sondra's name at trial, so she becomes "Miss X." In the novel Sondra's anonymity has a symbolic rather than a practical purpose. When he has the lawyers call Sondra "Miss X," Dreiser shows us that it is possible to understand her appeal to Clyde as long as one knows that she is a car-for-her-sixteenth-birthday type of girl. Her particular identity, like that of other characters, *see supra* text accompanying notes 98-101, is not central to understanding the events in which she has taken part.

125. DREISER, *supra* note 38, at 681.

blameless, they feel they must alter the facts of his story in order to argue for his legal innocence. If, as Belknap and Jephson believe, Clyde's plotting and subsequent failure to rescue Roberta make him just as guilty of first degree murder as if he had intended to kill her when he struck her with the camera,[126] then a defense and "the truth" are incompatible. The law does not allow the contention that a sane defendant who committed proscribed acts in the absence of external coercion is blameless nonetheless.[127] Thus, although Belknap and Jephson differ from Mason in having had Clyde tell them what "really happened,"[128] their obligation to provide him with a legal

---

126. *Id.* at 599. Dreiser gives the reader no reason to question this belief. Indeed, he encourages her to assume it is correct by describing Jephson as having unusually shrewd "mental and legal equipment." *Id.* at 598. Donald Pizer has suggested that Dreiser's opinion, as evidenced by Jephson's comment, that Clyde is legally guilty, was based on conversations with his lawyer friend Arthur Carter Hume. *See* PIZER, *supra* note 23, at 271.

The question of whether Clyde is, in fact, legally guilty is complex and must involve an analysis of his actions both before and after Roberta plunges into the water. Leo Katz, who acknowledges what happens is "shrouded in ambiguity by Dreiser's deliberately garbled prose," concentrates on the events leading up to Roberta's fall from the boat. *See* LEO KATZ, BAD ACTS AND GUILTY MINDS 202, 201-09 (1987). He argues these events are susceptible to at least three plausible interpretations, none of which offers a strong case for Clyde's guilt. As to whether Clyde's failure to rescue Roberta from the water after her plunge constitutes murder, Katz states the relevant questions are whether Clyde had a legal duty to save Roberta, and, if he did, whether he could have saved her if he had tried. *Id.* at 326 n.50.

Albert Levitt, who won the 1926 competition sponsored by Dreiser's publisher for the best essay entitled "Was Clyde Griffiths Guilty of Murder in the First Degree?", argues, for reasons similar to those offered in Katz's first interpretation, that Clyde's behavior before Roberta's plunge does not make him guilty of first degree murder. His intent to kill does not coexist with the acts of his that sent Roberta into the water. Albert Levitt, *Was Clyde Griffiths Guilty of Murder in the First Degree?* 4 (pamphlet distributed by Boni and Liverwright, 1926, on file at the Library of the University of Iowa, Iowa City, Iowa). Levitt goes on, unlike Katz, to consider the question of whether Clyde's omission to save Roberta is "culpable in law." He concludes that it is not according to general principles of the criminal law and the penal law of New York as it stood in 1926, because Clyde is not Roberta's parent, guardian, or husband. *Id.* at 5.

Even if Clyde is legally innocent, however, the general point that Dreiser makes in the context of Belknap's and Jephson's story—that the law will only countenance a limited number of determinist readings of defendants' actions—remains warranted.

127. *Cf.* Levitt, *supra* note 126, at 7 (affirming that "organized society" allows no defense to murder "based on weakness of the will, unless such weakness amounts to insanity").

128. Sensing Belknap's sympathy for him, Clyde has told him of his flight from Kansas City, his life in Lycurgus, all the details of his plot to drown Roberta, and how she came to die. Dreiser makes it clear that Belknap's sympathy, like Mason's resentment, is the result of his particular history and temperament. In his youth Belknap "himself had been trapped between two girls, with one of whom he was merely playing while being seriously in love with the other." DREISER, *supra*

defense leads them to fight him on his own sincere account.

Their first strategy in constructing Clyde's defense is to impose a legally excusing interpretation on it. Because the law tolerates seemingly exceptional, quasi-medical determinisms such as insanity,[129] they decide to relate the facts as Clyde has told them but to contend that "an illusion of grandeur aroused in Clyde by Sondra Finchley and the threatened disruption by Roberta of all his dreams and plans" caused a "brain storm," rendering Clyde temporarily insane.[130] They are, however, prevented from entering an insanity plea by the Lycurgus Griffiths, who are paying for Clyde's defense and who do not wish the stain of insanity to attach to "the Griffiths' blood and brain."[131]

Having had to reject the only legally acceptable means of arguing that Clyde should not be held responsible for plotting to kill Roberta and then letting her drown, Belknap and Jephson feel compelled to alter the plot of Clyde's story. They construct a narrative wherein Clyde does not plot to kill Roberta but, instead, takes her to the lake in order to tell her of his love for Sondra and thus convince her that he cannot marry her.[132] According to their story, once Clyde goes away with Roberta, spends two nights with her and sees how much she still loves him, he experiences a "change of heart" and decides to propose to her. Clyde takes Roberta out boating and offers to marry her while they are on the water. Roberta is so grateful for his proposal that she jumps up to come toward him. As she moves forward, her foot or dress catches, and she stumbles. Clyde, camera in hand, rises instinctively to try to catch her and stop her fall. As a result of their movement, the boat rocks and flips over,

---

note 38, at 592. When the girlfriend for whom he cared less became pregnant, Belknap avoided having to choose between marriage and permanent flight only through the efforts of his father, who sent the pregnant girl to another town with a thousand dollars and housing expenses. This experience, in conjunction with a well-trained mind and a wide experience of the world outside Cataraqui County, leads Belknap not to judge Clyde too harshly and, accordingly, to be inclined to see his behavior in determinist terms. *Id.* at 600. When Belknap tells Jephson of his first meeting with Clyde he states:

> "He could plot to kill one girl and maybe even did kill her, for all I know, after seducing her, but because he was being so sculled around by his grand ideas of this other girl, he didn't quite know what he was doing, really. Don't you see? You know how it is with some of these young fellows of his age, and especially when they've never had anything much to do with girls or money, and want to be something grand."

Id.

129. *See* KELMAN, GUIDE, *supra* note 6, at 91 (alluding to our preference for such determinisms over the sociological variety, which could have a more directly subversive impact on the justifiability of the criminal law).

130. DREISER, *supra* note 38, at 607.

131. *Id.*

132. *Id.* at 608-13.

and Clyde and Roberta fall into the water. Because the boat strikes Clyde as well as Roberta and makes him a little dizzy, by the time he becomes truly aware of her cries, she has drowned. At this point Clyde remembers his love for Sondra and slips away without calling any attention to the incident.[133]

The battle between Mason's narrative and the Belknap/Jephson account at Clyde's trial may be described as a fight within selective intentionalism as to whether Clyde fits into the category of those who have free will or those whose actions are predetermined. To cast the trial in this way is, however, to ignore its larger implications. Even a selective determinism—especially one whose primary selection mechanisms, youth and deprivation, designate a significant portion of the population—can undermine our faith in the general applicability of the intentionalist model.

Dreiser illustrates the way in which the intentionalist bias of the criminal law prevents us from recognizing the tension between unexceptional determinism and the justifiability of that bias. Legal categories ensure that the only determinist story on offer is one that does not account adequately for the evidence and that Clyde cannot testify to with force or conviction.[134] Moreover, by altering the plot of Clyde's story in this manner and rendering him innocent not only because he is the sort of person who cannot control his actions or desires but also because he did not do it, Belknap and Jephson obscure the subversive potential of unexceptional determinism.

An American Tragedy reveals not only legal suppression of the contradiction between faith in the generalizability of the intentionalist model and unexceptional determism but its existence in American culture as a whole. As I discuss above, Dreiser illustrates the way in which our practices of blame and punishment, and the legal categories designating determinist excuses as exceptional, rest on such a faith. And he provides evidence of the pull of unexceptional determinism that outweighs the jurors' failure to endorse the determinist description of Clyde.[135] Mason feels the need to emphasize Clyde's social

---

133. *Id.*

134. *See id.* at 692. At trial Clyde protested:

> "No! No! I never did plot to kill her, or any one," . . . clutching at the arms of his chair and seeking to be as emphatic as possible, since he had been instructed to do so. At the same time he arose in his seat and sought to look stern and convincing, although in his heart and mind was the crying knowledge that he had so plotted, and this it was that most weakened him at this moment—most painfully and horribly weakened him.

Id.

135. It is unclear from Dreiser's description of them whether the jurors are completely confident in intentionalism or whether, as farmers, clerks, and shopkeepers resentful of Clyde's upscale connections, they are simply unconvinced that he fits into the category of those whose actions are predetermined.

advantages in arguing for his responsibility for his actions; Belknap and Jephson buttress Clyde's defense with generalizations about what "we" cannot help; and Belknap, whom Dreiser describes as an example of a well-educated, sophisticated man, thinks of Clyde as the type of man who could not help acting as he did.[136] In American culture as a whole, then, Dreiser shows both the presence of the contradiction between our endorsement of unexceptional determinism and our faith in the general nature of the intentionalist model and the capacity of the law to obscure it.

CONCLUSION

The narratives of *An American Tragedy* present two different sorts of intentionalism/determinism conflicts. The overarching determinism of Dreiser's narrative conveys a picture of action that excludes the possibility of the selective intentionalism Mason evokes. Mason's selective intentionalism is not, however, inconsistent with the more limited determinism to which Belknap and Jephson are committed. The clash that occurs at trial is thus less between contradictory world views than between contradictory categorizations or descriptions of the defendant.

The conflict at trial is significant nonetheless because it illustrates how the intentionalist bias of legal discourse ensures that determinist description is only allowed in such ways as do not threaten the general practice of blaming. Clyde's lawyers cannot offer the argument that he was immature and deprived and therefore not responsible for his actions as a legal defense. The selective but unexceptional determinism implicit in such an argument would directly threaten the notion that it is justifiable for the legal system to consider free will the norm.

The space Dreiser's narrative inhabits is less severely constricted.[137] And the structural relationship among the stories in *An American Tragedy* privileges the full-blown determinist model of conduct embedded in the central narrative. As I noted earlier, for the reader Dreiser's narrative depicts what really happened, whereas Mason's account and the Belknap/Jephson tale are only stories told to a jury. In this way Dreiser ensures that the reader, unlike the jurors, will not be able to avoid feeling uneasy about the law's reliance on intentionalist assumptions.

Treating literary and legal narratives as cultural artifacts that have something to say about one another, as both this article and

---

136. *See supra* note 128 (description of Belknap's vision of Clyde).

137. Brook Thomas, among others, has noted that, although law and literature both reveal "the stories that a culture tells about itself," literature is capable of producing a wider range of narratives than is the law. *See* THOMAS, *supra* note 4, at 5.

An American Tragedy do, thus has a potentially disruptive effect. The novel's complete evocation of the determinist model of behavior can lead to uneasiness not only with the criminal law but also with the various moral frameworks on which we rely in daily life. Alan Gewirth has argued that the idea that agents "control or can control their behavior by their unforced choice while having knowledge of relevant circumstances,"[138] is a common feature of our various moralities.[139] Although Gewirth himself disposes of determinism,[140] the grinding ethos of *An American Tragedy* is discomfiting in part because it does not seem to give the reader that option.

The novel is also unsettling because, given our tendency to feel most comfortable when we believe our ideas rest on noncontingent foundations, it does not convince its reader that she may replace an old "truth," intentionalism, with a new one, determinism. In *An American Tragedy* Dreiser may encourage the reader to believe that human behavior is determined, but he also, perhaps unwittingly, teaches her to be suspicious of claims by narrative to describe a found, rather than created, world. By showing how the prosecution and the defense build their stories and by juxtaposing these stories with one another and with the central narrative of the novel, *An American Tragedy* illustrates the way in which a form of discourse that presents itself as the neutral depiction of facts rests on debatable theoretical perspectives. The reader's discomfort upon finishing the novel may thus increase as a result of her knowledge that *An American Tragedy* is itself a construction.

---

138. Alan Gewirth, *The Epistemology of Human Rights*, 1 SOC. PHIL. & POL'Y 1, 9 (1984).

139. *See id.* at 8.

140. ALAN GEWIRTH, REASON AND MORALITY 36-37 (1978).

RECENT DEVELOPMENTS

FAMILY LAW—TENNESSEE COURTS—RETROACTIVE ABOLITION OF THE COMMON LAW TORT OF CRIMINAL CONVERSATION

Hanover v. Ruch, 809 S.W.2d 893 (Tenn.), *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-525)

Sandra Hanover, the appellee's wife, was a patient of the appellant Robert M. Ruch, M.D., a gynecologist.[1] The appellant and Mrs. Hanover had been engaged in a sexual affair that began in late October of 1983 and continued for two years, ending in November of 1985.[2] Because of the affair, the appellee filed a divorce action against his wife, Sandra, to end their marriage of 31 years.[3] The appellee also filed a civil complaint against the appellant alleging three causes of action: criminal conversation, alienation of affections,[4] and medical malpractice.[5] At trial, the jury returned a verdict for the appellant on the alienation of affections and medical malpractice counts but found for the appellee on the claim of criminal conversation.[6] The jury assessed compensatory damages of $25,000 and punitive damages of $100,000.[7] The trial court approved the verdicts, and Dr. Ruch appealed.[8] The Tennessee Court of Appeals

---

1. Hanover v. Ruch, 809 S.W.2d 893, 893, *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-525).

2. *Id.* at 893. The Hanovers married in 1955. During their marriage, they had six children. Hanover v. Ruch, No. 58, 1990 WL 53276 (Tenn. Ct. App. May 1, 1990), *rev'd*, 809 S.W.2d 993 (Tenn.), *cert. denied*, 60 U.S.L.W. 3142 (1991) (No. 91-525). Mrs. Hanover, however, testified that there had been sexual dysfunction in their marriage, and that in their sex relationship Mr. Hanover was satisfied, but she was not. *Id.* Mrs. Hanover confided this to the appellant, and he invited her back to his office after visiting hours. *Id.*

3. *Hanover*, 809 S.W.2d at 893. The trial court granted a divorce on the ground of adultery. *Hanover*, 1990 WL 53276, at *1. Before the time of the affair, neither of the Hanovers had contemplated getting a divorce, and there is proof Mrs. Hanover had never had an extramarital affair other than this one. *Id.* at *5.

4. The Tennessee General Assembly abolished the tort of alienation of affections effective July 1, 1989, but the Legislature's action had no bearing on this lawsuit because the alleged alienation of affections occurred before the date of abolishment. *Hanover*, 1990 WL 53276, at *2; *see* TENN. CODE ANN. § 36-3-701 (1991) (originally enacted as 1989 Tenn. Pub. Acts 902).

5. *Hanover*, 809 S.W.2d at 893.

6. *Id.*

7. *Id.*

8. *Id.*

affirmed.[9] The Tennessee Supreme Court granted the appellant's application for appeal to determine whether the court should abolish the common law tort of criminal conversation.[10] On appeal to the Tennessee Supreme Court, *held*, reversed and dismissed. Because the reasons for the tort of criminal conversation no longer exist, and the public policy of the state is offended by criminal conversation actions, the tort of criminal conversation is abolished.[11] *Hanover v. Ruch*, 809 S.W.2d 893 (Tenn.), *cert. denied*, 60 U.S.L.W. 3342 (1991).

Section 39-13-508 of the Tennessee Code Annotated prospectively abolished the common law tort of criminal conversation.[12] Before this, criminal conversation was expressly recognized by the Tennessee Legislature as a tort for which a person could bring a cause of action.[13] *Hanover v. Ruch*[14] was filed before the legislative abolish-

---

9. *Id.* The Tennessee Court of Appeals recommended abolishment of the tort of criminal conversation but recognized it had no power to act and therefore affirmed the trial court. *Id.;* 1990 WL 53276, at *3-4; *see infra* notes 88-98 and accompanying text.

10. *Hanover*, 809 S.W.2d at 893.

11. *Id.* at 898. The court retroactively abolished the tort of criminal conversation. *Id.*

12. 1990 Tenn. Pub. Acts 773 (codified at TENN. CODE ANN. § 39-13-508 (1991)). Section 39-13-508 provides:

> (a) On or after January 1, 1991, no cause of action shall be maintained that is based upon the common law torts of seduction or criminal conversation, and such torts are hereby abolished.
>
> (b) Nothing in this section shall be construed as prohibiting a cause of action based upon sexual offenses, which offenses include, but are not limited to, those set out in this part.

TENN. CODE ANN. § 39-13-508.

Sections 1-4 of the 1990 Tenn. Pub. Acts state:

> SECTION 1. On or after the effective date of this act, no cause of action shall be maintained that is based upon the common law torts of seduction or criminal conversation and such torts are hereby abolished. Nothing in this section shall be construed as prohibiting a cause of action based upon a sexual offense which offenses shall include, but not be limited to, those set out in Tennessee Code Annotated, Title 39, Chapter 13, Part 5.
>
> SECTION 2. Tennessee Code Annotated, Section 28-3-104(a), is amended by deleting the words and punctuation "criminal conversation, seduction."
>
> SECTION 3. Tennessee Code Annotated, Section 20-1-106, is repealed in its entirety.
>
> SECTION 4. This act shall take effect on January 1, 1991, the public welfare requiring it, provided any action filed prior to such date may be maintained under the law in effect on the date of such filing.

1990 Tenn. Pub. Acts 773.

13. TENN. CODE ANN. § 28-3-104 (1980) (amended 1990). Before amendment in 1990, section 28-3-104 stated:

> (a) Actions for libel, for injuries to the person, false imprisonment, malicious prosecution, criminal conversation, seduction, breach of marriage promise brought under the federal civil rights statutes, and actions for

ment of criminal conversation.[15] During the decade before the suit was filed, however, the nationwide trend had been to abolish this particular tort.[16] The Tennessee Supreme Court faced the question whether it should follow the national trend and abolish the cause of action for the tort of criminal conversation.[17]

The Tennessee Court of Appeals has defined criminal conversation as "adulterous relations between the defendant and the spouse of the plaintiff."[18] The court found criminal conversation to be synonymous with adultery.[19] To recover damages for criminal conversation, the courts require a valid marriage between the spouses

---

statutory penalties shall be commenced within one (1) year after cause of action accrued.

(b) For the purpose of this section . . . no person shall be deprived of his right to maintain his cause of action until one (1) year from the date of his injury, and under no circumstances shall his cause of action be barred before he sustains an injury.

TENN. CODE ANN. § 28-3-104. The Legislature amended Section 28-3-104 to delete the words and punctuation "criminal conversation, seduction." *See* 1990 Tenn. Pub. Acts 773 § 2, *supra* note 12.

14. 809 S.W.2d 893 (Tenn.), *cert. denied*, 60 U.S.L.W. 3342 (1991)(No. 91-523).

15. *Id.* at 893, 895.

16. *Id.* Many states used text similar to Tennessee's statute to abolish the tort of criminal conversation. *See* ALA. CODE § 6-5-331 (1975); CAL. CIV. CODE § 43.5 (West 1982); COLO. REV. STAT. ANN. § 13-20-202 (West 1989); CONN. GEN. STAT. ANN. § 52-572f (West Supp. 1991); DEL. CODE ANN. tit. 10, § 3924 (1984); D.C. CODE ANN. § 16-923 (1989); FLA. STAT. ANN. § 771.01 (West 1986); GA. CODE ANN. § 51-1-17 (Michie 1982); IND. CODE ANN. § 34-4-4-1 (Burns 1986); MICH. COMP. LAWS ANN. § 600.2901 (West 1986); MINN. STAT. ANN. § 553.02 (West 1988); NEV. REV. STAT. § 41.380 (1987); N.J. STAT. ANN. § 2A:23-1 (West 1987); N.Y. CIV. RIGHTS LAW § 80-a (McKinney 1976); OHIO REV. CODE ANN. § 2305.29 (Anderson 1991); OKLA. STAT. ANN. tit. 76, § 8.1 (West 1987); OR. REV. STAT. § 30.850 (1988); TEX. FAM. CODE ANN. § 4.05 (West Supp. 1984); VT. STAT. ANN. tit. 15, § 1001 (1989); VA. CODE ANN. § 8.01-220 (Michie 1984); WYO. STAT. § 1-23-101 (1977). Maryland and Montana abolished alienation of affections but not criminal conversation. *See* MD. CTS. & JUD. PROC. CODE ANN. § 5-301 (1989); MONT. CODE ANN. § 27-1-601 (1991); *see also* W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 124, at 929-31 (5th ed. 1984) [hereinafter THE LAW OF TORTS]; Marshall L. Davidson, III, Comment, *Stealing Love in Tennessee: The Thief Goes Free*, 56 TENN. L. REV. 629, 656 (1989).

17. *Hanover*, 809 S.W.2d at 893. The court stated that because the action had no place in contemporary society, the social harm it creates outweighed the benefits, and the Legislature declared it to be against public policy, it must abolish criminal conversation. *Id.* at 898.

18. *Hanover*, 1990 WL 53276, at *1; *accord* Rheudasil v. Clower, 270 S.W.2d 345, 346 (Tenn. 1954); Darnell v. McNichols, 122 S.W.2d 808, 811 (Tenn. Ct. App. 1938); *see* 41 AM. JUR. 2D *Husband and Wife* § 478 (1968); THE LAW OF TORTS, *supra* note 16, at 917; RESTATEMENT (SECOND) OF TORTS § 685 cmt. d (1976); Davidson, *supra* note 16, at 652.

19. *Hanover*, 1990 WL 53276, at *1.

and sexual intercourse between the defendant and plaintiff's spouse.[20] The plaintiff need not prove alienation of the spouse's affections. Alienation of affections, however, will be regarded as a matter of aggravation.[21] "Consent of the participating spouse is not a defense."[22] Consequently, the only defense to an action for criminal conversation is consent of the plaintiff.[23] Criminal conversation is similar to strict liability, for "[r]ecovery is assured upon proof of the marriage between plaintiff and his spouse and an act of adultery occurring between the defendant and [the] plaintiff's spouse during the marriage."[24]

Punishment for adultery was first exhibited by the Teuton tribes.[25] Most tribes fixed a penalty that the adulterer had to pay the husband.[26] Much later, the Anglo-Saxons established the action of criminal conversation as part of the early English common law.[27] This early common law tort action originated from actions against third parties by the master for the enticement away of his servants, thereby depriving him of their services.[28] Because the station of a wife under

---

20. *Id.*; THE LAW OF TORTS, *supra* note 16, at 917; RESTATEMENT (SECOND) OF TORTS § 685 (1976); *see also* Davidson, *supra* note 16 at 652.

21. The alienation of affections tort is distinct from the action for criminal conversation. *Darnell*, 122 S.W.2d at 810. In *Darnell*, the plaintiff sued the defendant for enticing the plaintiff's wife into leaving him, seeking $10,000 in damages. *Id.* The trial court held for the plaintiff and awarded him $3,000. *Id.* at 809. The defendant appealed, averring that the plaintiff individually should have pleaded criminal conversation and alienation of affections. *Id.* at 810-11. In affirming the trial court, the court of appeals wrote that it is possible for criminal conversation and alienation of affections to exist one without the other. *Id.* at 810. The court further stated:

> While an action for alienation of affections and one for criminal conversation are both founded on the injury to the right of consortium they are generally recognized as essentially different. The gravamen or gist of the action where it is for criminal conversation is the adulterous intercourse, and the alienation of affections thereby resulting is regarded as merely a matter of aggravation, whereas the gravamen in the other case is the alienation of the affections with malice or improper motives.

Id.; *see* THE LAW OF TORTS, *supra* note 16, at 921.

22. *Hanover*, 1990 WL 53276 at *1; *see* THE LAW OF TORTS, *supra* note 16, at 921; 41 AM. JUR. 2D *Husband and Wife* § 478 (1968).

23. *Hanover*, 1990 WL 53276, at *1; Stepp v. Black, 14 Tenn. App. 153, 160 (1931); *see* THE LAW OF TORTS, *supra* note 16, at 921; RESTATEMENT (SECOND) OF TORTS § 687 (1976).

24. *Hanover*, 1990 WL 53276 at *1; *see* THE LAW OF TORTS, *supra* note 16, at 921.

25. *Hanover*, 809 S.W.2d at 894; *see* Jacob Lippman, Note, *The Breakdown of Consortium*, 30 COLUM. L. REV. 651, 654 (1930). In some instances, the husband was allowed to kill his wife's lover. *Id.* at 654-55. This punishment was later reduced to allowing the husband to merely emasculate the adulterer. *Id.* at 655.

26. *See* Lippman, *supra* note 25, at 655. "[A]mong the Anglo-Saxons the amount [of damages] depended on the station in life of the husband." *Id.*

27. *Id.*

28. *Id.* at 653.

early common law was that of a servant of the husband, an action was available to include the loss of her services.[29] Other purposes underlying an action for criminal conversation included vindicating the husband's property rights in the wife's person and punishing the defendant for defiling the plaintiff's marriage bed, besmirching family honor, and placing the legitimacy of the children in doubt.[30]

In seventeenth century England, the criminal conversation action became very important because it was necessary to obtain a judgment against the lover as a condition precedent to procuring a divorce through a private act of Parliament.[31] With the establishment of a divorce court having the right to grant a decree of divorce for the adultery of the wife, the action for criminal conversation was no longer needed.[32] The tort of criminal conversation was abolished by statute in England in 1857.[33]

The common law tort of criminal conversation has existed in the United States since the 1800s.[34] At early common law, suits for criminal conversation could be brought only by the husband.[35] Because of the wife's disability of coverture,[36] she had no right to bring an action for criminal conversation.[37] This changed with the passage of the Married Women's Emancipation Act.[38] After its enactment, a wife could sue for criminal conversation.[39]

In the past few decades, however, the national trend has been to abolish the tort of criminal conversation by statute.[40] Many states also have abolished the tort by judicial decision.[41] In *Koestler v.*

---

29. *Id.*; THE LAW OF TORTS, *supra* note 16, at 916; Lippman, *supra* note 25, at 653.

30. *See generally* THE LAW OF TORTS, *supra* note 16, at 917-18; Lippman, *supra* note 25, at 655.

31. Lippman, *supra* note 25, at 659.

32. *Id.* at 659-60.

33. *Id.*; 21 Vict. Ch. 85, § 59 (1857).

34. *See, e.g.*, Bigaouette v. Paulet, 134 Mass. 123 (1883).

35. *See id.*; 14 TENN. JUR. *Husband and Wife* § 31 (1984); Lippman, *supra* note 25, at 655.

36. Coverture is the condition or state of a married woman. BLACK'S LAW DICTIONARY 366 (6th ed. 1990).

37. *See* Scates v. Nailing, 268 S.W.2d 561, 563 (Tenn. 1954); TENN. JUR., *supra* note 35, at 352 n.17.

38. This law is codified at TENN. CODE ANN. § 36-3-504 (1991).

39. *See Scates*, 268 S.W.2d at 563.

40. *See supra* note 16.

41. Pickering v. Pickering, 434 N.W.2d 758, 763 (S.D. 1989); Destafano v. Grabrian, 763 P.2d 275, 279 (Colo. 1988) (holding that the tort of criminal conversation no longer existed but allowing an action for breach of fiduciary duty against a priest who was acting as a marriage counselor and committed adultery with a client); Harrington v. Pages, 440 So. 2d 521, 522 (Fla. Dist. Ct. App. 1983); Hyman v. Moldovan, 305 S.E.2d 648 (Ga. Ct. App. 1983); Bilyk v. Chicago Transit Auth., 531 N.E.2d 1, 7 (Ill. 1988); Gasper v. Lighthouse, Inc., 533 A.2d 1358, 1361

Pollard,[42] the Supreme Court of Wisconsin held a state statute abolishing the tort of criminal conversation, as well as public policy, barred a husband's claim for damages.[43] Plaintiff had filed an action for emotional distress against the biological father of his wife's child.[44] The plaintiff's wife became pregnant with the child during an adulterous relationship with the defendant while she was married to the plaintiff.[45] The trial court dismissed the action.[46] The court of appeals affirmed.[47] On appeal to the Supreme Court of Wisconsin, the court affirmed the lower court's decision,[48] reasoning the action was essentially one of criminal conversation, which had been barred by the Legislature.[49] The court stated the public policy of the state was well served by the abolition of the tort of criminal conversation.[50] One of the reasons people of the state benefit from the abolition of the tort is that it is often misused in a vindictive manner.[51]

(Md. 1987) (dismissing a suit for criminal conversation against the marriage counselor of the plaintiff's spouse because the Legislature had previously abolished the tort); Cotton v. Kambly, 300 N.W.2d 627, 628 (Mich. Ct. App. 1980) (holding that the statute abolishing the tort of criminal conversation precluded judgment on that course of action but did not bar an action for medical malpractice); Larson v. Dunn, 449 N.W.2d 751, 756 (Minn. Ct. App. 1990); Feldman v. Feldman, 480 A.2d 34, 36 (N.H. 1984) (abolishing the tort of criminal conversation because the tort diminished human dignity); Zaragoza v. Capriola, 492 A.2d 698, 702 (N.J. Super. Ch. Div. 1985); Cannon v. Miller, 322 S.E.2d 780, 804 (N.C. Ct. App. 1984), *vacated*, 327 S.E.2d 888 (1985); Hardy v. VerMeulen, 512 N.E.2d 626, 628 (Ohio 1987); Lund v. Caple, 675 P.2d 226, 231 (Wash. 1984) (abolishing the tort citing the wrongful motives of the plaintiff as a rationalization); Brown v. Thomas, 379 N.W.2d 868, 873 (Wis. 1985) (citing the abolition of the "heart balm" statutes). For a discussion of *Cannon* see *infra* notes 52-64 and accompanying text.

42. 471 N.W.2d 7 (Wis. 1991).

43. *Id.* at 9.

44. *Id.* at 8.

45. *Id.* The plaintiff discovered the true father of the child when the defendant revealed the child's paternity. *Id.* This occurred after the plaintiff already had bonded with the child. *Id.*

46. 471 N.W.2d at 9. The trial court stated that the plaintiff had failed to state a claim upon which relief may be granted and reasoned that the Legislature had intended to abolish plaintiff's cause of action when it abolished claims for criminal conversation and alienation of affections. *Id.*

47. *Id.* at 7. The court of appeals certified appeal to the Supreme Court of Wisconsin. *Id.*

48. *Id.*

49. *Id.* at 9. The court noted the plaintiff's complaint was drafted in an attempt to avoid WIS. STAT. § 768.01 (1990), which would bar an action for criminal conversation. *Id.* The statute states: "Actions for breach of promise, alienation of affection and criminal conversation abolished. All causes of action for breaches to marry, alienation of affections and criminal conversation are hereby abolished, except that this section shall not apply to contracts now existing or to causes of action which heretofore accrued." WIS. STAT. § 768.01 (1990).

50. *Koestler*, 471 N.W.2d at 12.

51. *Id.* at 11.

In *Cannon v. Miller*,[52] the North Carolina Court of Appeals held the torts of alienation of affections and criminal conversation were abolished.[53] The plaintiff brought an action against the defendant seeking damages for alienation of affections and criminal conversation.[54] Plaintiff alleged the defendant, an attorney, persuaded the plaintiff's wife to have sexual relations with him.[55] The defendant claimed he did not have sex with the plaintiff's wife until after the plaintiff and his wife were separated.[56] The trial court granted the defendant's motion for summary judgment, and the plaintiff appealed.[57] The North Carolina Court of Appeals affirmed the trial court.[58] In its holding, the court attempted to abolish the torts of criminal conversation and alienation of affections.[59] The court gave many reasons for the abolishment of these torts.[60] First, the torts attract disproportionate publicity.[61] Second, the torts have outlived their purposes and no longer have a place in modern society.[62] Third, plaintiffs frequently abuse these actions.[63] Finally, the court stated the primary justification behind the judicial abolition of the tort of criminal conversation was the lack of logically valid defenses on the merits.[64]

---

52. 322 S.E.2d 780 (N.C. Ct. App. 1984), *rev'd*, 327 S.E.2d 888 (N.C. 1985)).

53. *Cannon*, 322 S.E.2d at 804. The Supreme Court of North Carolina, however, reversed the holding stating that the appellate court did not have the power to abolish a cause of action. Cannon v. Miller, 327 S.E.2d 888 (N.C. 1985).

54. *Cannon*, 322 S.E.2d at 783. Plaintiff sought both actual and punitive damages for a total amount of $250,000. *Id.*

55. *Id.* at 783. The plaintiff alleged that the defendant affected the will of his wife in such a way as to transfer her love from the plaintiff to the defendant. *Id.* The plaintiff and his wife divorced soon after the affair. *Id.*

56. *Id.* The defendant had proof the plaintiff and his wife were not happily married, they had been separated for some time, and his acquaintance with the plaintiff's wife began after the separation. *Id.*

57. *Id.* at 784.

58. *Id.* at 804. The court wrote a lengthy opinion detailing the history of the torts of criminal conversation and alienation of affections. *See id.* at 780. This opinion was cited as persuasive authority by the Supreme Court of Tennessee in *Hanover*, 809 S.W.2d at 897.

59. *Cannon*, 322 S.E.2d at 804.

60. *See id.* at 794-804.

61. *Id.* at 794. Because these torts suggest sexual misbehavior, people's emotional and moral indignation often prevail over considerations of private or public injury in the assessment of damages. *Id.*

62. *Cannon*, 322 S.E.2d at 794. The torts were established to protect a man's property from others. *See* Lippman, *supra* note 25, at 658. Now that women are no longer regarded as their husbands' property, awarding damages to the husbands for their wives adulterous relationships is no longer appropriate. *See* THE LAW OF TORTS, *supra* note 16.

63. *Cannon*, 322 S.E.2d at 796. The action provides opportunities for blackmail. *Id.*; *see infra* note 102. Plaintiffs also use these actions in vindictive manners to force settlements or embarrass the defendants. *Koestler*, 471 N.W.2d at 11.

64. *Cannon*, 322 S.E.2d at 796. To recover damages for criminal conversation,

In Tennessee, criminal conversation means "adulterous relations between the defendant and the spouse of the plaintiff."[65] To recover for criminal conversation, the plaintiff must prove only a valid marriage between the spouses and sexual intercourse between defendant and plaintiff's spouse during the time of the marriage.[66] The amount of damages recoverable is primarily a matter for the jury to decide.[67]

The tort of criminal conversation was first recognized in Tennessee by the court of appeals in *Stepp v. Black*.[68] Black, the plaintiff, sued the defendant Stepp for criminal conversation.[69] The trial court awarded plaintiff $8,000 in damages[70] and defendant appealed.[71] The Tennessee Court of Appeals affirmed the trial court's decision but modified the judgment.[72] The court approved the action of criminal conversation by stating, "it follows as a matter of law that the plaintiff is entitled to recover a substantial amount as compensation for loss of consortium, impairment of his wife's affections, loss of her services, and his own mental anguish and even physical suffering and impairment . . . ."[73]

In *Rheudasil v. Clower*,[74] the Tennessee Supreme Court affirmed the existence of an action in tort for criminal conversation.[75] Accordingly, the Tennessee Legislature also has recognized this tort.[76]

---

the plaintiff must show only a valid marriage between the spouses and the occurrence of sexual intercourse between the defendant and the plaintiff's spouse. THE LAW OF TORTS, *supra* note 16, at 917; RESTATEMENT (SECOND) OF TORTS § 685 (1976); Davidson, *supra* note 16, at 652. The only defense to the action is connivance of the plaintiff. THE LAW OF TORTS, *supra* note 16, at 921.

65. Rheudasil v. Clower, 270 S.W.2d 345, 346 (Tenn. 1954); *see* 14 Tenn. App. 153; 14 TENN. JUR. *Husband and Wife* § 31 (1984).

66. *Hanover*, 1990 WL 53276, at *1.

67. Sweeny v. Carter, 137 S.W.2d 892, 896 (Tenn. Ct. App. 1939).

68. 14 Tenn. App. 153 (1931).

69. *Id.* at 155. Plaintiff and his wife were married in 1919, and lived together until 1929. *Id.* At that time they separated but lived as tenants on the same farm. *Id.* The defendant, a neighbor, admitted having sexual intercourse with the plaintiff's wife. *Id.* at 157.

70. *Id.* at 155.

71. *Id.*

72. *Id.* The court suggested a remittitur of $3,000 to bring the award down to $5,000. *Id.* at 172. The court stated that if the remittitur was not accepted the entire judgment would be reversed and the cause remanded for a new trial. *Id.* at 172-73.

73. *Stepp* at 172. The court stated the defendant was a man of fine character who attended church but "fell by temptation." *Id.*

74. 270 S.W.2d 345 (Tenn. 1954).

75. *Id.* The court, in a brief opinion, held because an action for criminal conversation was subjected to a one year statute of limitation period, an action for the alienation of affections was also limited by this one year period. *Id.* at 346.

76. *See supra* note 12.

Before the decision in *Hanover v. Ruch*,[77] many states had abolished by legislative act the tort of criminal conversation.[78] The courts of many states also have abolished this common law tort.[79] In fact, the Tennessee Legislature passed a bill in 1990 abolishing this action.[80]

Tennessee courts have suggested abolishing the tort of criminal conversation.[81] The courts refer to decisions of other states in abolishing this action.[82] The rationalizations upon which these courts rely for abolishing this specific tort are many.[83]

In *Lentz v. Baker*,[84] the Eastern Section of the Tennessee Court of Appeals discussed the need to abolish the common law tort of alienation of affections.[85] In 1987, the plaintiff in *Lentz* brought an

---

77. 809 S.W.2d 893 (1991), *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-525).

78. *See supra* note 16. In 1935, the Illinois Legislature enacted a statute popularly known as the "Heart Balm Law" making it unlawful to file an action for alienation of affections, criminal conversation, and breach of contract to marry. Daily v. Parker, 61 F. Supp. 701, 702 (N.D. Ill. 1945). The court held that making it unlawful to file the action did not abolish the tort of criminal conversation. *Id.* at 703.

79. *See supra* note 41.

80. *See*, 1990 Tenn. Pub. Acts 773, *supra* note 12.

81. *Hanover*, 1990 WL 53276 at *3; Lentz v. Baker, 792 S.W.2d 71, 77 (Tenn. Ct. App. 1989).

82. *Hanover*, 1990 WL 53276, at *3. The court found persuasive the argument of the North Carolina Court of Appeals for the abolition of the tort of criminal conversation in Cannon v. Miller, 322 S.E.2d 780 (N.C. Ct. App. 1984), *vacated*, 327 S.E.2d 888 (N.C. 1985). *Hanover*, 1990 WL 53276 at *2. Although the *Cannon* court held the tort of criminal conversation was abolished, the North Carolina Supreme Court subsequently reversed and remanded for trial, holding the intermediate appellate court had no authority to abolish a common law tort. Cannon v. Miller, 327 S.E.2d 888 (N.C. 1985).

83. *See* Fundermann v. Mickelson, 304 N.W.2d 790 (Iowa 1981); Wymann v. Wallace, 615 P.2d 452 (Wash. 1980); Simpson v. Dufresne, No. 23793-7, 1991 WL 81651 (Tenn. Ct. App. May 21, 1991); *Hanover*, 1990 WL 53276 at *3; *Lentz*, 792 S.W.2d at 71; Cannon v. Miller, 322 S.E.2d 780, 803 (N.C. Ct. App. 1984); *see also Davidson, supra* note 16, at 656.

In Wyman v. Wallace, 615 P.2d 452 (Wash. 1980), the Supreme Court of Washington abolished the tort, citing the following reasons for abolition:

> (1) The underlying assumption for preserving marital harmony is erroneous; (2) The judicial process is not sufficiently capable of policing the often vicious out-of-court settlements; (3) The opportunity for blackmail is great since the mere bringing of an action could ruin the defendant's reputation; (4) There are no helpful standards for assessing damages; and (5) The successful plaintiff succeeds in compelling what appears to be a forced sale of the spouse's affections.

Id. at 455.

The actions for alienation of affections and criminal conversation apparently never existed in the state of Louisiana. *See* Moulin v. Monteleone, 115 So. 447, 448 (La. 1927); Ohlhausen v. Brown, 372 So. 2d 787, 788 (La. Ct. App. 1979).

84. 792 S.W.2d 71 (Tenn. Ct. App. 1989).

85. *Id.* at 74-77.

action against his wife's preacher for alienation of affections.[86] The plaintiff and his wife divorced in 1984, at which time the plaintiff learned that the preacher and his wife had been having sexual relations.[87] The trial court found that the defendant was liable but did not award any damages.[88] The plaintiff appealed.[89] The Tennessee Court of Appeals affirmed the decision of the trial court.[90] While affirming the lower court's holding, the court wrote that the action of alienation of affections "does not protect marriage or the family . . . and the harm it causes far outweighs any reason for its continuance."[91] The court also stated these actions diminish human dignity, demean the participants, and do not prevent human misconduct.[92]

On appeal, the Western Section of the Tennessee Court of Appeals in *Hanover*[93] advocated the abolition of the tort of criminal conversation.[94] The court wrote that the torts of criminal conversation and alienation of affections[95] have become "fertile ground for blackmail and extortion by means of manufactured suits brought by plaintiffs with vindictive motives."[96] The court also stated that the primary justification for the abolition of the tort of criminal conversation was the lack of logically valid defenses on the merits.[97] According to the court, another fundamental flaw of the tort is

---

86. *Id.* at 72. The plaintiff and his wife married in 1976. *Id.* at 71. Plaintiff was a boilermaker and experienced lengthy periods of unemployment. *Id.* When employed, he worked long hours and was frequently out of town. *Id.* As a result of the long hours, the absences, and the lack of money, the marriage was under stress. *Id.* Sexual incompatibility and plaintiff's physical abuse of the wife were also problems in the marriage. *Id.*

87. *Id.* During the pendency of the divorce, the plaintiff's wife admitted she had been having a sexual relationship with the defendant, a relationship that had begun in 1981. *Id.*

88. *Id.* at 71.

89. *Id.* The plaintiff alleged on appeal that the jury's finding of liability without awarding damages was inconsistent. *Id.*

90. *Id.* at 76.

91. *Id.* The court indicated it affirmed the trial court's decision solely because an appellate court lacks the power to abolish the actions. *Id.*

92. *Id.* at 76 (quoting *Wyman*, 549 P.2d at 74).

93. No. 58, 1990 WL 53276 (Tenn. Ct. App. May 1, 1990).

94. *Hanover*, 1990 WL 53276, at *3. Here, as in *Lentz*, 792 S.W.2d at 71, the court was powerless to abolish the common law action. *Id.*

95. The General Assembly of Tennessee previously had abolished the tort of alienation of affections. 1989 Tenn. Pub. Acts 902.

96. *Hanover*, 1990 WL 53276, at *2.

97. *Id.* The court explained that "apart from the fact that the action for criminal conversation . . . was anachronistic because the antiquated common law reasoning of the wife's inferiority which lay 'behind the stripping of the defendant of all defenses to an action of criminal conversation, save the plaintiff's consent, the cause of action no longer merits endorsement.'" *Id.* (quoting *Cannon*, 322 S.E.2d at 796).

allowing recovery of damages where the marriage remains unaffected.[98]

The Tennessee Court of Appeals in *Hanover*[99] took into account the adverse results an action for criminal conversation causes the defendants in suggesting the tort's abolition.[100] These actions are easily abused by plaintiffs seeking to discredit defendants.[101] The potential damage to reputations and the threat to sue can easily become, in effect, schemes of extortion and blackmail.[102] For example, a party in a divorce may use the threat of a suit to force the other party into a disproportionate settlement.[103]

The Tennessee Supreme Court in *Hanover v. Ruch*[104] agreed with the Western and Eastern Sections of the Court of Appeals and held the tort of criminal conversation was obsolete and should be abolished in Tennessee.[105] The court stated the reasons for the cause of action no longer existed, and that criminal conversation actions contravened the public policy of the state, as expressed by the Legislature.[106]

The court abolished the tort of criminal conversation retroactively.[107] The Tennessee Supreme Court determined the court had the power to do this to a common law tort.[108] The Legislature is without authority to enact a retrospective law affecting vested substantive rights.[109] Where a statute alters the common law and results in deprivation of a valuable common law right, such statute cannot be applied retroactively.[110] In contrast, retroactive application of judicial changes in the common law is a unique feature of common law

---

98. *Id.*

99. No. 58, 1990 WL 53276 (Tenn. Ct. App. May 1, 1990), *rev'd*, 809 S.W.2d 893 (Tenn.), *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-525).

100. *Id.* at *3. The reasons for the abolition of the tort are similar to those found in *Cannon*, 322 S.E.2d at 796.

101. *Hanover*, 1990 WL 53276, at *3.

102. *Id.* Professor Prosser has commented that the actions for criminal conversation and alienation of affections have "afforded a fertile field for blackmail and extortion by means of manufactured suits in which the threat of publicity is used to force a settlement." THE LAW OF TORTS, *supra* note 16, § 124, at 929.

103. *Hanover*, 1990 WL 53276, at *3.

104. 809 S.W.2d 893 (Tenn. 1991), *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-525).

105. *Id.* at 898.

106. *Id.* The court recognized that it held contrary to the precedent set by earlier Tennessee courts. *Id.* the court stated, however, that "the law changes when necessary to serve the people." *Id.*

107. *See id.* By reversing the appellate court and holding for the defendant, the Tennessee Supreme Court abolished the tort of criminal conversation retroactively. *Id.*

108. *Id.* at 896.

109. *Id.* The Tennessee Constitution states: "No retrospective laws. - That no retrospective law, or law impairing the obligations of contracts, shall be made." TENN. CONST. art. I, § 20.

110. *Hanover*, 809 S.W.2d at 896; *see* Massey v. Sullivan County, 464 S.W.2d 548 (Tenn. 1971).

adjudication.[111] Where the Legislature has not acted, the Tennessee courts have abolished obsolete common law doctrines.[112] Because the Legislature had determined that the public policy of the state was served by the abolition of the tort of criminal conversation,[113] and precedent gave the Tennessee Supreme Court the power to retroactively abolish obsolete common law doctrine,[114] the court decided it was able to abolish the tort of criminal conversation retroactively.[115]

Abolishing the action retroactively could leave the door open for courts to interpret any legislative act that abolishes a common law tort as being abolished retroactively. This judicial legislation would lead to chaos, for the long-standing doctrine of *stare decisis*[116] could be disregarded when a court determines a common law tort to be obsolete. In *Hanover*,[117] however, the court limited its holding to the common law tort of criminal conversation.[118] The court held only that prospective legislation abolishing a common law tort does not deprive the court of the power to modify common law doctrines in previously filed cases.[119] According to the *Hanover*[120] court, prospective legislation is a factor for the court's consideration but is not dispositive of retrospective common law adjudication.[121]

With the holding in *Hanover*,[122] Tennessee has joined a number of states that have abolished judicially the common law tort of criminal conversation.[123] The manner in which the court held in this case could aid future courts in abolishing other common law doctrines that do more harm than good, for it affirms the power of the courts to abolish these outdated actions.[124]

Constitutional problems can arise if the courts try to use their power to abolish any tort action.[125] The Tennessee Constitution has

---

111. *Hanover*, 809 S.W.2d at 896; *see* Davis v. Davis, 657 S.W.2d 753, 759 (Tenn. 1983).

112. *Hanover*, 809 S.W.2d at 896; *see, e.g.*, *Davis*, 657 S.W.2d at 758 (abolishing interspousal tort immunity doctrine); Kilbourne v. Hanzelik, 648 S.W.2d 932, 934 (Tenn. 1983) (abolishing discriminatory rule denying liability of wife for support of husband).

113. *See* 1990 Pub. Acts 773, *supra* note 12.

114. *See supra* note 112.

115. *Hanover*, 809 S.W.2d at 897.

116. *Stare decisis* means: "To abide by, or adhere to, decided cases." BLACK'S LEGAL DICTIONARY 1261 (5th ed. 1976).

117. 809 S.W.2d 893.

118. *Id.* at 896.

119. *Id.*

120. *Id.* at 893.

121. *Id.* at 896 n.5.

122. *Id.* at 893.

123. *Id.*; *see supra* note 41.

124. The Tennessee Court of Appeals already has held in accordance with the *Hanover* decision without any dissent. Simpson v. Dufresne, No. 23793-7, 1991 WL 81651 (Tenn. Ct. App. May 21, 1991).

125. *See* TENN. CONST. art. II.

provisions regulating what branch has the power to make the laws of the state.[126] The courts are prohibited by the constitution from modifying acts passed by the Legislature.[127] Therefore, the courts have the power to abolish solely the common law torts, which were created by the judiciary at their inception.[128]

Hanover v. Ruch[129] represents the trend to abolish common law torts that are outdated and have no value to present day society.[130] The tort of criminal conversation, which has been subject to considerable abuse,[131] had been abolished prospectively by the Legislature[132] and now has been abolished retrospectively by the Tennessee Supreme Court.[133] This decision benefits state residents, for lawsuits that have no just reason for existence can subject people to unnecessary and embarrassing litigation. This author believes it is the duty of the courts to protect all citizens from unjust and unfair uses of the courts. If one of these practices includes using a law that is very easily abused for the sole purpose of damaging another, the court has the duty to act. In *Hanover*,[134] the court acted by retroactively abolishing the common law tort claim of criminal conversation.[135] The court has the power to do this[136] and, in this case, wielded that power appropriately.

GARRETT E. ASHER

---

126. Tennessee Constitution article II reads:

Sec. 1. Division of powers.—The powers of the Government shall be divided into three distinct departments: the Legislative, Executive, and Judicial.

Sec. 2. Limitation of powers.—No person or persons belonging to one of these departments shall exercise any of the powers properly belonging to either of the others, except in the cases herein directed or permitted.

TENN. CONST. art. II, §§ 1-2.

127. *See supra* note 126.

128. *See Hanover*, 809 S.W.2d at 896.

129. 809 S.W.2d 893 (1991), *cert. denied*, 60 U.S.L.W. 3342 (1991) (No. 91-528).

130. *See* Davidson, *supra* note 15.

131. *See, e.g., id., supra* note 15.

132. *See supra* note 15.

133. *Hanover*, 809 S.W.2d at 898.

134. 809 S.W.2d at 893.

135. *Id.*

136. *See supra* note 112.

TORTS—PERSONAL LIABILITY OF GOVERNMENT EMPLOYEES—FEDERAL EMPLOYEES LIABILITY REFORM AND TORT COMPENSATION ACT

United States v. Smith, 111 S. Ct. 1180 (1991)

Dominique Smith was born at a United States Army hospital in Vicenza, Italy, in 1982, allegedly with massive brain damage.[1] Five years later Dominique's parents, as plaintiffs, filed suit in the United States District Court for the Central District of California against Dr. William Marshall, who served as attending physician at the birth.[2] Because the defendant was a federal employee, the United States intervened pursuant to the Gonzalez Act[3] and requested the government be substituted as the sole defendant.[4] The United States further argued that after the substitution, the action should be governed by the Federal Tort Claims Act (FTCA),[5] and the suit, consequently, should be dismissed pursuant to the FTCA exception precluding

---

1. United States v. Smith, 111 S. Ct. 1180, 1183 (1991). At the time of Dominique's birth, his father, an Army sergeant, was stationed in Italy. *Id.*

2. *Id.* The complaint, which based jurisdiction on diversity of citizenship, alleged the defendant's negligence caused Dominique's injuries. *Id.*

3. Gonzalez Act, Pub. L. No. 94-464, § 1(a), 90 Stat. 1985 (1976) (codified at 10 U.S.C. § 1089 (1988)). The Gonzalez Act provides that the Federal Tort Claims Act is the exclusive remedy for injuries arising from malpractice by medical personnel acting within the scope of their duties for the Department of Defense. *Id.* § 1089(a). The Act requires the Attorney General to defend or settle any medical malpractice actions against military medical personnel. *Id.* § 1089(b)-(d). Finally, the Act authorizes the appropriate agency head to provide liability insurance for military medical personnel when in situations where the Federal Tort Claims Act does not apply. *Id.* § 1089(f). Section 1089(a) states in pertinent part:

> The remedy against the United States provided by sections 1346(b) and 2672 of title 28 for damages for personal injury, including death, caused by the negligent or wrongful act or omission of any physician . . . of the armed forces . . . while acting within the scope of his duties or employment therein or therefore shall hereafter be exclusive of any other civil action or proceeding by reason of the same subject matter against such physician . . . whose act or omission gave rise to such action or proceeding.

10 U.S.C. § 1089(a) (1988).

4. *Smith*, 111 S. Ct. at 1183.

5. *Id.* Under the FTCA, the United States consents to be sued in certain situations for money damages. Primarily these situations are for loss of property, personal injury, or death, caused by a negligent or wrongful act or omission of any federal employee acting within the scope of his or her employment. 28 U.S.C. § 1346(b) (1988).

recovery for any claim arising in a foreign country.[6] The district court agreed with the government, substituted the United States as the sole defendant, and dismissed the action.[7] While the plaintiffs' appeal was pending, Congress passed the Federal Employees Liability Reform and Tort Compensation Act of 1988 (Liability Reform Act),[8] which amended the FTCA.[9] Eschewing reliance on the Gonzalez Act, the United States contended in the court of appeals the Liability Reform Act required substitution of the government as the sole defendant.[10] The United States Court of Appeals for the Ninth Circuit reversed the district court's decision, however, and found neither the Liability Reform Act nor the Gonzalez Act required substitution of the United States as the sole defendant.[11] Further, the Ninth Circuit held neither act immunized Dr. Marshall from liability.[12] On certiorari to the United States Supreme Court, *held*, reversed. The Liability Reform Act confers immunity on federal employees for common law

---

6. *Smith*, 111 S. Ct. at 1183. The FTCA does not apply to "[a]ny claim arising in a foreign country." 28 U.S.C. § 2680(k)(1988). The Supreme Court has interpreted the phrase "foreign country," as used in the FTCA, to mean "territory subject to the sovereignty of another nation." United States v. Spelar, 338 U.S. 217, 219 (1949).

7. *Smith*, 111 S. Ct. at 1183. The district court also found the plaintiffs had failed to file an administrative claim within the time limits prescribed by law, which was a prerequisite to suit under the FTCA. *Id.* at 1183 n.2. According to section 2401(b) of title 28, an administrative claim must be presented to the agency, in writing, within two years of accrual of the cause of action, and suit must be filed within six months of denial of the claim by the agency. 28 U.S.C. § 2401(b) (1988).

8. Pub. L. No. 100-694, 102 Stat. 4563 (codified at 16 U.S.C. § 831c-2 and 28 U.S.C. §§ 1 note, 2671, 2671 notes, 2674, 2679, 2679 note (1988)). The Liability Reform Act provides absolute immunity to federal employees for torts committed within the scope of their employment and makes an FTCA action against the United States the exclusive remedy for such torts. 28 U.S.C. § 2679(b)(1) (1988). The Act applies to "all claims, civil actions, and proceedings pending on, or filed on or after, the date of enactment of this Act." 28 U.S.C. § 2679 note (1988). In *Smith*, neither party contested the applicability of the Liability Reform Act to the case. *Smith*, 111 S. Ct. at 1184 n.4.

9. *Smith*, 111 S. Ct. at 1183.

10. *Id.* at 1184. The United States argued because Dr. Marshall's alleged malpractice occurred within the scope of his employment, the plaintiffs' exclusive remedy was under the FTCA. *Id.* Further, the United States contended, because the FTCA bars recovery for injuries occurring overseas, the district court appropriately dismissed the complaint. *Id.*

11. Smith v. Marshall, 885 F.2d 650, 651 (9th Cir. 1989). Even though the United States relied solely on the Liability Reform Act before the court of appeals, the Ninth Circuit addressed the applicability of the Gonzalez Act "[i]n view of an apparent split among the circuits . . . ." *Id.* at 651-52.

12. *Id.* at 653, 655. Because the FTCA does not apply to a claim arising in a foreign country, the Ninth Circuit reasoned the Liability Reform Act was not applicable to the plaintiff's suit and, thus, did not immunize the military physician. *Id.* at 655.

torts committed while acting within the scope of their employment when an FTCA exception precludes recovery against the United States. *United States v. Smith*, 111 S. Ct. 1180 (1991).

In 1988, the Liability Reform Act amended the FTCA and limited the relief available to tort victims.[13] For a person injured by a federal employee acting within the scope of employment, the Liability Reform Act provides that an FTCA remedy against the United States is exclusive of any other civil action against the employee.[14] Pursuant to the Liability Reform Act, an injured party may sue only the United States; the injured party is statutorily barred from pursuing a cause of action against the individual employee.[15] The courts of appeals, however, have disagreed on the applicability of the provisions of the Liability Reform Act to the FTCA in situations involving one of the FTCA's specific exceptions.[16] The First, Fifth, and Tenth

---

13. 28 U.S.C. § 2679(b)(1) (1988).

14. *Id.* This section states:

> The remedy against the United States provided by sections 1346(b) and 2672 of this title for injury or loss of property, or personal injury or death arising or resulting from the negligent or wrongful act or omission of any employee of the Government while acting within the scope of his office or employment is exclusive of any other civil action or proceeding for money damages by reason of the same subject matter against the employee whose act or omission gave rise to the claim or against the estate of such employee. Any other civil action or proceeding for money damages arising out of or relating to the same subject matter against the employee or the employee's estate is precluded without regard to when the act or omission occurred.

Id.

15. *Id.*

16. Pursuant to 28 U.S.C. § 2680, the FTCA does not apply to the following:

> (a) Any claim based upon an act or omission of an employee of the Government, exercising due care, in the execution of a statute or regulation, whether or not such statute or regulation be valid, or based upon the exercise or performance or the failure to exercise or perform a discretionary function or duty on the part of a federal agency or an employee of the Government, whether or not the discretion involved be abused.
>
> (b) Any claim arising out of the loss, miscarriage, or negligent transmission of letters or postal matter.
>
> (c) Any claim arising in respect of the assessment or collection of any tax or customs duty, or the detention of any goods or merchandise by any officer of customs or excise or any other law-enforcement officer.
>
> (d) Any claim for which a remedy is provided by sections 741-752, 781-790 of Title 46, relating to claims or suits in admiralty against the United States.
>
> (e) Any claim arising out of an act or omission of any employee of the Government in administering the provisions of sections 1-31 of Title 50, Appendix.
>
> (f) Any claim for damages caused by the imposition or establishment of a quarantine by the United States.
>
> [(g) Repealed.]
>
> (h) Any claim arising out of assault, battery, false imprisonment, false

Circuits found the Liability Reform Act applies where an FTCA exception precludes a liability action against the United States.[17] In contrast, the Eleventh and Ninth Circuits held the Act, as an amendment to the FTCA, only applies in situations where the FTCA itself applies.[18] In June 1990, the United States Supreme Court granted certiorari to review the Ninth Circuit's holding in *Smith v. Marshall*.[19] The question presented in *Smith* was whether an injured person whom the FTCA foreign-country exception precluded from suing the United States may seek recovery from the particular federal employee who caused the injury.[20]

The law always has been well settled: the United States, as sovereign, is immune from suit unless the government has consented to be sued.[21] As for federal employees, the judicial system usually provided the employee with immunity from personal tort liability when the employee was acting within the scope of employment.[22] As early as 1875, the Supreme Court held agents acting for the government were protected by its authority while acting within the scope of their powers.[23] Twenty-one years later in *Spalding v. Vilas*,[24] the Supreme Court explicitly stated that conduct by a head of an exec-

---

arrest, malicious prosecution, abuse of process, libel, slander, misrepresentation, deceit, or interference with contract rights . . . .

(i) Any claim for damages caused by the fiscal operations of the Treasury or by the regulation of the monetary system.

(j) Any claim arising out of the combatant activities of the military or naval forces, or the Coast Guard, during time of war.

(k) Any claim arising in a foreign country.

(l) Any claim arising from the activities of the Tennessee Valley Authority.

(m) Any claim arising from the activities of the Panama Canal Company.

(n) Any claim arising from the activities of a Federal land bank, a Federal intermediate credit bank, or a bank for cooperatives.

28 U.S.C. § 2680 (1988).

17. Nasuti v. Scannell, 906 F.2d 802 (1st Cir. 1990) (assault and battery exception, 28 U.S.C. § 2680(h)); Mitchell v. Carlson, 896 F.2d 128 (5th Cir. 1990) (assault and battery exception, 28 U.S.C. § 2680(h)); Aviles v. Lutz, 887 F.2d 1046 (10th Cir. 1989) (defamation and interference with contract rights exceptions, 28 U.S.C. § 2680(h)).

18. Smith v Marshall, 885 F.2d 650 (9th Cir. 1989) (foreign country exception); Newman v. Soballe, 871 F.2d 969 (11th Cir. 1989) (foreign country exception).

19. 885 F.2d 650 (9th Cir. 1989), *cert. granted sub nom.*, United States v. Smith, 110 S. Ct. 2617 (1990).

20. United States v. Smith, 111 S. Ct. 1180, 1183 (1991).

21. United States v. Sherwood, 312 U.S. 584, 586 (1941). *See, e.g.*, United States v. Lee, 106 U.S. 196, 205-07 (1882); United States v. Thompson, 98 U.S. 486, 489 (1878); United States v. Clarke, 33 U.S. (8 Pet.) 151, 154 (1834).

22. *See infra* notes 23-25 and accompanying text.

23. Lamar v. Browne, 92 U.S. 187 (1875).

24. 161 U.S. 483 (1896).

utive department "cannot be made the foundation of a suit against him personally for damages" when the executive is acting within his authority, regardless of any personal motives.[25]

The situation changed in 1946 with the enactment of the FTCA.[26] According to the terms of the FTCA, the United States consented to be sued for the negligent or wrongful acts or omissions of employees acting within the scope of their employment.[27] The FTCA as originally enacted did not expressly prohibit suits against individual employees.[28] The FTCA provided, however, that if a tort victim chose to sue the United States, the victim was precluded from subsequently suing the individual employee.[29] Typically, a cause of action against a federal employee[30] was unsuccessful because the courts often held employees were protected by immunity, particularly when the act leading to the injury involved a discretionary function.[31]

Discretionary functions took on an even greater significance in January 1988 when the Supreme Court announced its decision in *Westfall v. Erwin*.[32] Resolving a dispute among the courts of appeals,[33] the Supreme Court held in order for an employee to be

---

25. *Id.* at 499.

26. Federal Tort Claims Act, ch. 753, title IV, 60 Stat. 842 (1946) (codified as amended at 28 U.S.C. §§ 1291, 1346(b)-(c), 1402(b), 2401(b), 2402, 2411, 2412(c), 2671-2680 (1988)).

27. 28 U.S.C. § 1346(b) (1988). Under the FTCA, the United States is liable "in the same manner and to the same extent as a private individual under like circumstances . . . ." *Id.* § 2674. The courts have consistently held the United States may define the terms and conditions under which it may be sued. *E.g.*, Honda v. Clark, 386 U.S. 484, 501 (1967); Soriano v. United States, 352 U.S. 270, 276 (1957); United States v. Sherwood, 312 U.S. 584, 586 (1941).

28. 28 U.S.C. § 2679(b) (1982) (amended 1988).

29. Section 2676 of title 28 provides "[t]he judgment in an action under section 1346(b) of this title shall constitute a complete bar to any action by the claimant, by reason of the same subject matter, against the employee of the government whose act or omission gave rise to the claim." 28 U.S.C. § 2676 (1988).

30. By suing the employee, a tort victim can avoid the terms and conditions established by the FTCA. For example, the FTCA provides that the victim must present an administrative claim to the agency in writing within two years of accrual of the cause of action. 28 U.S.C. § 2401(b) (1988). Also, an FTCA action is not tried by a jury. 28 U.S.C. § 2402 (1988). In addition, attorney fees in an FTCA action are limited to 25% of any judgment. 28 U.S.C. § 2678 (1988). The United States is also not liable for interest prior to judgment or punitive damages. 28 U.S.C. § 2674 (1988).

31. *See* Barr v. Matteo, 360 U.S. 564, 574-75 (1959) (absolute privilege extends to federal employees acting within the scope of duty); Howard v. Lyons, 360 U.S. 593, 597 (1959) (claim of absolute privilege judged by federal standards formulated by the courts in absence of legislative action by Congress); Dalehite v. United States, 346 U.S. 15, 33 (1953) (employees performing discretionary functions cannot be sued).

32. 484 U.S. 292 (1988).

33. *Compare* General Elec. Co. v. United States, 813 F.2d 1273, 1276-77

absolutely immune from state-law tort liability, the employee's act must be discretionary, as well as within the scope of employment.[34] This decision significantly broadened the scope of federal employees' liability.[35] Appealing for legislative direction, the Court commented "Congress is in the best position to provide guidance for the complex and often highly empirical inquiry into whether absolute immunity is warranted in a particular context. Legislated standards governing the immunity of federal employees involved in state-law tort actions would be useful."[36]

Congress responded to the Supreme Court's request with the enactment of the Liability Reform Act.[37] Unhappy with the ramifications of *Westfall*,[38] Congress observed that the decision "dramatically changed the law which governs the personal tort liability of Federal employees."[39] Thus, the Liability Reform Act was passed to "provide immunity for Federal employees from personal liability for common law torts committed within the scope of their employment."[40]

(4th Cir. 1978) *and* Poolman v. Nelson, 802 F.2d 304, 307 (8th Cir. 1986) (federal employees are absolutely immune from tort suits for conduct within the scope of employment) *with* Johns v. Pettibone Corp., 769 F.2d 724, 728 (11th Cir. 1985) *and* Araujo v. Welch, 742 F.2d 802, 806 (3d Cir. 1984) (federal employees are absolutely immune from tort suits for discretionary acts within the scope of employment).

34. *Westfall*, 484 U.S. at 295.

35. *See infra* notes 38-39 and accompanying text.

36. *Westfall*, 484 U.S. at 300.

37. *See supra* note 8.

38. Congressman Wolf commented "it is unconscionable to ask Government employees, who have no reasonable way of knowing whether they are protected when they act, to make decisions or take actions capable of spawning tort lawsuits." 134 CONG. REC. H4719 (daily ed. June 27, 1988) (statement of Cong. Wolf). Similarly, Senator Grassley noted:

> As a result of Westfall, we are now faced with an immediate crisis of personal liability exposure for the entire Federal workforce—more than 3 million persons in all three branches of government . . . . Virtually every one of these employees—and particularly rank-and-file civil servants—now face the possibility of being required to defend a lawsuit in which his or her personal fortune is a [sic] stake—even when the actions complained of were clearly official duties.
>
> . . . .
>
> The prospect of years of personal liability litigation against the Federal workforce not only has a devastating impact on individual civil servants' pocketbooks, credit ratings, and morale, but will severely inhibit the ability of many agencies to carry out their mission.

134 CONG. REC. S15599 (daily ed. Oct. 12, 1988) (statement of Sen. Grassley).

39. H.R. REP. No. 100-700, 100th Cong., 2d Sess. 2, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5946. *See also* 134 CONG. REC. H4719 (daily ed. June 27, 1988) (statement of Cong. Moorhead) ("decision has placed all Federal employees in a state of uncertainty with regard to personal liability").

40. H.R. REP. No. 100-700, 100th Cong., 2d Sess. 2, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5945.

The Liability Reform Act provides that individuals harmed by the acts of federal employees[41] will be able to sue the United States for damages.[42] Such suits are exclusive of any other civil action against the employee whose act gave rise to the claim.[43] Describing the effect of the Act, Congress stated "suits against Federal employees are precluded even where the United States has a defense which prevents an actual recovery. Thus, any claim against the government that is precluded by the exceptions set forth in Section 2680 of Title 28, U.S.C. also is precluded against an employee . . . ."[44] The provisions of the Liability Reform Act do not extend to (1) constitutional torts[45] or (2) violations of a federal statute under which an action against an individual is authorized.[46]

In 1989, federal courts of appeals began interpreting the Liability Reform Act in suits against federal employees. *Newman v. Soballe*,[47] decided by the United States Court of Appeals for the Eleventh Circuit in April 1989, involved an incident of alleged medical malpractice by a military physician in Japan.[48] Although the Eleventh Circuit reached its decision that a military physician could be sued individually for alleged malpractice occurring overseas based on the Gonzalez Act,[49] the court also discussed the applicability of the Liability Reform Act.[50] Stating the Liability Reform Act "is expressly designed to immunize from personal tort liability those federal employees not protected by other immunity statutes," the Eleventh

---

41. The Liability Reform Act defines "employee of the government" to include "officers or employees of any federal agency, members of the military or naval forces of the United States . . . ." 28 U.S.C. § 2671 (1988).

42. 28 U.S.C. § 2679(b)(1) (1988).

43. *Id.* Section 2679(b)(1) states: "[t]he remedy against the United States provided by [the FTCA] . . . is exclusive of any other civil action . . . against the employee . . . ." 28 U.S.C. § 2679(b)(1) (1988). This section represents the codification of section 5 of the Liability Reform Act.

44. H.R. REP. No. 100-700, 100th Cong., 2d Sess. 6, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5950.

45. 28 U.S.C. § 2679(2)(A) (1988). *See infra* note 109. *See also* Bivens v. Six Unknown Agents of the Fed. Bureau of Narcotics, 403 U.S. 388 (1971) (federal employees may be sued individually for constitutional torts).

46. 28 U.S.C. § 2679(2)(B) (1988). *See infra* note 109.

47. 871 F.2d 969 (11th Cir. 1989).

48. *Id.* After dismissal in federal court for lack of diversity or other federal jurisdiction, the plaintiffs refiled the complaint in state court. *Id.* The United States requested removal to federal court, substitution of the United States as the sole defendant, and dismissal pursuant to the FTCA exclusion of claims arising overseas. *Id.* at 969-70. Because the Liability Reform Act had not yet been enacted, the United States requests were based on the Gonzalez Act. *Id. See also supra* note 3. The district court granted the motion and dismissed the complaint for lack of subject matter jurisdiction. *Newman*, 871 F.2d at 970.

49. *Id.* at 970-71.

50. *Id.* at 971. The complaint in *Newman* was filed before Congress enacted the Liability Reform Act. *Id.*

Circuit found the defendant was not "among those federal employees affected" by the later legislation.[51] Because the defendant was covered under the Gonzalez Act he was, according to the Eleventh Circuit, excluded from the protection of the Liability Reform Act.[52]

A similar interpretation of the Liability Reform Act was subsequently adopted by the United States Court of Appeals for the Ninth Circuit in *Smith v. Marshall*.[53] The Ninth Circuit held neither the Gonzalez Act nor the Liability Reform Act immunized a military physician practicing abroad from medical malpractice claims.[54] The court found that because the plaintiffs had no remedy against the United States under the FTCA,[55] the Liability Reform Act was "inapplicable" and did not provide the physician with immunity.[56]

In reaching this conclusion, the Ninth Circuit reviewed the Liability Reform Act's language and its legislative history. The court focused on section 9 of the Act, which specifically extends the protection of the Liability Reform Act to Tennessee Valley Authority (TVA) employees.[57] The court reasoned if Congress had intended to extend immunity coverage under the Liability Reform Act to federal employees for tort claims arising in foreign countries, Congress expressly would have done so as it did for TVA employees.[58] Further, the court found the legislative history of the Act to be contradictory.[59] The court noted Congress explained that the exclusive remedy provision of section 5 means the United States is substituted as the solely permissible defendant in all common law tort actions even when the

---

51. *Id.*

52. *Id.* The Eleventh Circuit determined the Liability Reform Act only applies in situations where the FTCA applies. *Id.* Because the plaintiff's claim arose in a foreign country, the court reasoned the FTCA did not apply. *Id.*

53. 885 F.2d 650 (9th Cir. 1989).

54. *Id.* at 651.

55. The Ninth Circuit's decision was based upon the foreign country exception to the FTCA that exempts the United States from liability in claims arising in a foreign country. *See* 28 U.S.C. § 2680(k) (1988).

56. *Smith*, 885 F.2d at 654-55.

57. As codified, section 9 of the Liability Reform Act states:

> An action against the Tennessee Valley Authority for injury or loss of property, or personal injury or death arising or resulting from the negligent or wrongful act or omission of any employee of the Tennessee Valley Authority while acting within the scope of this office or employment is exclusive of any other civil action or proceeding by reason of the same subject matter against the employee or his estate whose act or omission gave rise to the claim. Any other civil action or proceeding arising out of or relating to the same subject matter against the employee or his estate is precluded without regard to when the act or omission occurred.

16 U.S.C. § 831c-2(a)(1) (1988) (footnote omitted).

58. *Smith*, 885 F.2d at 655.

59. *Id.*

FTCA exceptions preclude a cause of action under the Act.[60] But, the court stated, Congress also said the Liability Reform Act would not diminish any legal right held by an individual.[61] Because a grant of immunity to the military physician would be tantamount to a denial of a legal right conferred by the Gonzalez Act to the plaintiffs, the court relied on what it termed consistent statutory language to deny immunity to the physician.[62]

The United States Court of Appeals for the Tenth Circuit subsequently refused to interpret the Liability Reform Act as inapplicable where FTCA exceptions preclude recovery.[63] *Aviles v. Lutz*[64] is a suit by a former federal employee, alleging defamation and tortious interference with employment rights by a supervisor.[65] *Aviles* addresses another exception to the FTCA that precludes actions for libel, slander, or interference with contract rights.[66] The Tenth Circuit held, in contrast to the decisions by the Ninth and Eleventh Circuits,[67] when an FTCA exception renders the FTCA inapplicable to a tort victim's claim, the victim still cannot pursue an action against the individual employee.[68] To reach that result the Tenth Circuit relied on the plain language of the Liability Reform Act and its legislative history.[69]

The Tenth Circuit noted section 2679(d)(4) of title 28 plainly states that upon the Attorney General's certification that a federal employee was acting within the scope of employment, any civil action against such an employee shall proceed under the FTCA, subject to the limitations and exceptions of the FTCA.[70] The court further

---

60. *Id. See* H.R. REP. No. 100-700, 100th Cong., 2d Sess. 6, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5950.

61. *Smith*, 885 F.2d at 655. *See* H.R. REP. No. 100-700, 100th Cong., 2d Sess. 7, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5951. When the Liability Reform Act was under consideration by the House of Representatives, Congressman Wolf commented, "[i]n no way does this measure infringe or diminish any legal rights of the individual." 134 CONG. REC. H4719 (daily ed. June 17, 1988) (statement of Cong. Wolf).

62. *Smith*, 885 F.2d at 656.

63. Aviles v. Lutz, 887 F.2d 1046 (10th Cir. 1989).

64. *Id.*

65. *Id.* at 1047. The district court dismissed the complaint in *Aviles*, holding the suit was barred under the doctrines of sovereign immunity, res judicata, and absolute immunity. *Id.*

66. 28 U.S.C. § 2680(h) (1988). *See supra* note 16.

67. *Smith*, 885 F.2d 650 (9th Cir. 1989); *Newman*, 871 F.2d 969 (11th Cir. 1989). *See supra* notes 47, 53 and accompanying text.

68. *Aviles*, 887 F.2d at 1049.

69. *Id.* at 1048-49.

70. *Id.* at 1049. Section 2679(d) is the codification of section 6 of the Liability Reform Act. It states, in part, that "[u]pon certification, any action . . . shall proceed in the same manner as any action against the United States filed pursuant to section 1346(b) of this title [FTCA] and shall be subject to the limitations and exceptions applicable to those actions. 28 U.S.C. § 2679(d)(4) (1988). For other certification provisions, see *infra* note 80.

found Congress' mandatory language within this provision does not permit a challenge to certification.[71] As additional evidence of legislative intent, the court quoted the House report on the Liability Reform Act, which plainly states any claim against the United States precluded by FTCA exceptions is also precluded against an individual employee.[72] In a case such as *Aviles*, where an FTCA exception precludes recovery, the court concluded "Congress had not waived the sovereign immunity of the United States and its employees."[73] The court, therefore, held plaintiffs did not have a cause of action against either the United States or the individual employee.[74]

The United States Court of Appeals for the Fifth Circuit likewise has interpreted the Liability Reform Act to apply where an FTCA exception precluded a cause of action.[75] In *Mitchell v. Carlson*,[76] an alleged tort victim sued a federal employee for assault and battery.[77] The Fifth Circuit ruled the district court should have dismissed the suit for lack of jurisdiction without remanding it to state court.[78] Commenting on the clear mandate of the exclusive remedy provision in section 5 of the Liability Reform Act,[79] the Fifth Circuit held an action against a federal employee who has been certified[80] as acting

---

71. *Aviles*, 887 F.2d at 1049.

72. *Id.* (quoting H.R. REP. No. 100-700, 100th Cong., 2d Sess. 6, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5950).

73. *Id.* at 1050.

74. *Id.* at 1049-50.

75. Mitchell v. Carlson, 896 F.2d 128 (5th Cir. 1990).

76. *Id.*

77. *Id.* at 130. The plaintiff filed suit in state court. Upon removal to federal court, the district court substituted the United States as the sole defendant and dismissed the action pursuant to the FTCA exclusion of claims arising from assault and battery. In the same order, the district court also resubstituted the employee as the defendant and remanded the case to state court. *Id.*

78. *Id.* at 134.

79. *Id. See supra* note 43.

80. Following are some of the certification provisions contained in the Liability Reform Act:

> (d)(1) Upon certification by the Attorney General that the defendant employee was acting within the scope of his office or employment at the time of the incident out of which the claim arose, any civil action or proceeding commenced upon such claim in a United States district court shall be deemed an action against the United States under the provisions of this title and all references thereto, and the United States shall be substituted as the party defendant.
>
> (d)(2) Upon certification by the Attorney General . . . any civil action or proceeding commenced upon such claim in a State court shall be removed without bond at any time before trial by the Attorney General to the district court of the United States . . . . This certification of the Attorney General shall conclusively establish scope of office or employment for purposes of removal.

28 U.S.C. § 2679(d)(1)-(2) (Supp. 1991). *See supra* note 70.

in the scope of employment must proceed exclusively against the United States under the FTCA.[81]

Rejecting the Ninth Circuit's characterization of the Act's legislative history as "internally inconsistent," the Fifth Circuit commented "[i]solated language found scattered throughout the legislative history is insufficient persuasion that Congress intended to frustrate the very purpose of the [Liability Reform Act], to protect its employees from the distraction and burden of litigation based upon their employment activities."[82] The Fifth Circuit stated that the Ninth Circuit ignored the general language of section 5, which refers to an "exclusive" remedy and specifically precludes any other civil action against the employee.[83] The Fifth Circuit further observed that Congress intended to leave a plaintiff such as the one in *Mitchell* without a remedy because the FTCA had no remedy provisions for assault and battery.[84] "Assault and battery are specifically excepted. Under the [Liability Reform Act], Mitchell cannot look elsewhere for a remedy."[85]

The United States Court of Appeals for the First Circuit next joined the Tenth and Fifth Circuits in applying the Liability Reform Act to situations where the FTCA precludes recovery.[86] In *Nasuti v. Scannell*,[87] the First Circuit held a federal employee sued for assault has an absolute right to personal immunity when certified by the Attorney General to have been acting within the scope of employment unless a federal court determines the employee, contrary to certification, was acting outside the scope of employment when the tort occurred.[88] The court stated the Liability Reform Act bars any claim

---

81. *Mitchell*, 896 F.2d at 134.

82. *Id.* at 136.

83. *Id.*

84. *Id.* at 134. *See* 28 U.S.C. § 2680(h) (1988 & Supp. 1991).

85. *Mitchell*, 896 F.2d at 135.

86. Nasuti v. Scannell, 906 F.2d 802 (1st Cir. 1990).

87. *Id. Nasuti* was before the First Circuit twice, once before and once after the enactment of the Liability Reform Act. Nasuti v. Scannell, 906 F.2d 802 (1st Cir. 1990) (*Nasuti*); Nasuti v. Scannell, 792 F.2d 264 (1st Cir. 1986) (*Nasuti I*). The court of appeals dismissed *Nasuti I* for lack of appellate jurisdiction. *Id.* at 803.

88. *Nasuti*, 906 F.2d at 803, 810. In *Aviles*, the Tenth Circuit stated the Attorney General's certification as to scope of employment could not be challenged by the court. *See supra* note 71 and accompanying text. Whereas in *Nasuti*, the First Circuit explained certification conclusively establishes scope of employment for removal purposes only and does not abolish "the district court's normal power to determine the scope question finally where necessary to resolve disputed rights and the court's own subject-matter jurisdiction." *Nasuti*, 906 F.2d at 808. *See supra* note 80 for text of § 2679(d)(2) (conclusiveness of the Attorney General's certification).

Section 2679(d)(3) explicitly provides for certification by the court when the Attorney General has refused to certify that an employee was acting within the scope of employment, and the employee petitions the court for such a finding. 28 U.S.C. § 2679(d)(3) (1988).

against a federal employee acting within the scope of employment.[89] To support this conclusion, the First Circuit noted Congress intended to protect employee immunity by providing Attorney General certification of employment is conclusive and by eliminating the remand provision of the Driver's Act.[90] The court found nothing in the Liability Reform Act or its legislative history to suggest Congress ever intended for an employee to lose immunity from suit simply because the plaintiff cannot recover against the United States under the FTCA.[91]

On March 20, 1991, the United States Supreme Court decided *United States v. Smith*[92] and thereby resolved the dispute among the circuits that had resulted in uncertainty concerning the liability of individual federal employees for common-law torts arising in situations exempt from FTCA remedies. In this 8-to-1 decision the Court held the Liability Reform Act confers immunity on federal employees for common-law torts committed while the employee is acting within the scope of employment and where an FTCA exception precludes recovery against the United States.[93] As a result, the plaintiffs in *Smith* were unable to seek damages for injuries allegedly caused by a military physician working for the United States in a military hospital in Italy.[94]

Writing for the majority, Justice Marshall reasoned the Liability Reform Act, which designates the FTCA as the exclusive remedy for common-law torts committed by federal employees, precludes any alternative method of recovery against such employees even where the FTCA does not provide a remedy against the government.[95] In reaching this decision, Justice Marshall applied well settled principles of statutory construction and explicitly rejected the statutory interpretation advanced by the lone dissenter, Justice Stevens, and by the Ninth Circuit in the decision below.[96]

At issue in *Smith* was the effect of the Liability Reform Act on the personal liability of federal employees acting within the scope of

---

89. *Nasuti*, 906 F.2d at 814.

90. *Id.* at 809 (referring to 28 U.S.C. § 2679(d)(2) (1988)). The pre-1988 version of section 2679(d) codified the Driver's Act, which provided a case removed to federal court that would be remanded to state court upon a determination by the federal district court that the case was one in which a remedy against the United States under the FTCA was not available. 28 U.S.C. § 2679(d) (1982), *amended by* 28 U.S.C. § 2679(d)(2) (1988).

91. *Nasuti*, 906 F.2d at 809.

92. *Smith*, 111 S. Ct. 1180 (1991).

93. *Id.* at 1184-85. The Supreme Court did not address the portion of the Ninth Circuit's ruling that denied Dr. Marshall immunity under the Gonzalez Act because the United States did not raise the issue on appeal. *Id.* at 1184 n.6.

94. *Id.* at 1183.

95. *Id.* at 1185.

96. *Id.* at 1185-95; Smith v. Marshall, 885 F.2d 650 (9th Cir. 1989).

their employment in situations where recovery against the United States was precluded by the FTCA.[97] The specific question addressed was whether an Army medical doctor could be sued for medical malpractice allegedly committed in Italy when no FTCA remedy was available because of the Act's exception for claims arising in a foreign country.[98]

The majority opinion in *Smith* focuses on section 5 of the Liability Reform Act, which provides the remedy against the United States under the FTCA "is exclusive of any other civil action or proceeding for money damages . . . against the employee."[99] Following the approach taken by the courts of appeals, Justice Marshall looked to the section's plain meaning, as supported by its legislative history.[100] To Justice Marshall and seven other justices the meaning of section 5 was clear: in common-law tort actions, the exclusive remedy is an action against the United States under the FTCA, and any action against the employee is precluded.[101] The Court found support for this construction in the "limitations and exceptions" language of section 6 of the Liability Reform Act[102] and the express preservations of employee liability in section 5.[103]

Section 6 provides for certification by the Attorney General that the defendant employee was acting within the scope of employment.[104] Upon certification, the action is to proceed in the "same manner as any action against the United States filed pursuant to" the FTCA and is "subject to the limitations and exceptions applicable to those actions."[105] The Supreme Court emphasized one of these limitations and exceptions is section 2680(k) of title 28, which bars liability for torts "arising in a foreign country."[106] Adopting the rationale of the Tenth Circuit in *Aviles*,[107] Justice Marshall was convinced Congress, by its use of the "limitations and exceptions" language, fully realized a plaintiff's recovery may be completely foreclosed.[108]

---

97. *Smith*, 111 S. Ct. at 1183.

98. *Id.* *See* 28 U.S.C. § 2680(k) (1988).

99. *Smith*, 111 S. Ct. at 1185. *See supra* note 43.

100. *Id.* at 1185.

101. *Id.*

102. The "limitations and exceptions" phrase of section 6 is codified at 28 U.S.C. § 2679(d)(4) (1988). *See supra* note 70.

103. *Smith*, 111 S. Ct. at 1185. The express preservations of employee liability are found at 28 U.S.C. § 2679(b)(2) (1990). *See infra* note 109.

104. *Smith*, 111 S. Ct. at 1185. *See* 28 U.S.C. § 2679(d)(1)-(2) (1988). *See also supra* note 80.

105. *Smith*, 111 S. Ct. at 1185 (quoting 28 U.S.C. § 2679(d)(4) (1988)). *See supra* note 70.

106. *Smith*, 111 S. Ct. at 1185 (referring to 28 U.S.C. § 2680(k) (1988)). *See supra* note 16.

107. *See supra* notes 71-73 and accompanying text.

108. *Smith*, 111 S. Ct. at 1185.

The Supreme Court next examined the provisions of section 5, which expressly preserve employee liability in two areas.[109] Relying on principles of statutory construction applied in *Andrus v. Glover Construction Company*,[110] the Supreme Court stated additional exceptions cannot be implied when Congress expressly lists specific exceptions.[111] Thus, under section 5, the only situations in which an individual federal employee may be sued based upon employment-related activity are ones involving either a constitutional tort or a violation of a federal law that specifically authorizes a suit against a federal employee.[112] Neither situation, however, existed in *Smith*.[113]

Justice Marshall rejected the plaintiffs' argument that their claim was covered under the exception in section 5 for violations of federal law.[114] The plaintiffs alleged an action for malpractice against a military doctor practicing abroad would involve a violation of the Gonzalez Act.[115] Justice Marshall reasoned "[n]othing in the Gonzalez Act imposes any obligations or duties of care on military physicians," and therefore, a physician who acts negligently under state or foreign law does not "violate" the Gonzalez Act.[116]

Justice Marshall also relied on the legislative history of the Liability Reform Act to further support his construction of sections 5 and 6.[117] The pertinent House report explains the exclusive remedy provision of section 5 by stating that the section provides for the substitution of the "United States as the *solely* permissible defendant in *all* common law tort actions against federal employees who acted in the scope of employment."[118] The report further states "any claim

---

109. *Id.* Section 2679(b)(2) provides:

> Paragraph (1) [reference to § 2679(b)(1)] does not extend or apply to a civil action against an employee of the Government—
>
> (A) which is brought for a violation of the Constitution of the United States, or
>
> (B) which is brought for a violation of a statute of the United States under which such action against an individual is otherwise authorized.

28 U.S.C. § 2679(2) (1988).

110. 446 U.S. 608 (1980).

111. *Id.* at 616-17.

112. *See supra* note 109.

113. *See Smith*, 111 S. Ct. at 1189 (alleged medical malpractice is not a violation of the Gonzalez Act).

114. *Id.* at 1189 (discussing § 2679(b)(2)(B)). *See supra* note 109.

115. *Smith*, 111 S. Ct. at 1189. *See supra* note 93 (Supreme Court did not address whether the Gonzalez Act denied immunity to Dr. Marshall).

116. *Smith*, 111 S. Ct. at 1189. The purpose of the Gonzalez Act as stated by Congress is to immunize certain medical personnel from liability for acts committed within the scope of their employment. S. REP. No. 94-1264, 94th Cong., 2d Sess. 1, *reprinted in* 1976 U.S.C.C.A.N. 4443, 4443.

117. *Smith*, 111 S. Ct. at 1185 n.9.

118. H.R. REP. No. 100-700, 100th Cong., 2d Sess. 6, *reprinted in* 1988 U.S.C.C.A.N. 5945, 5950 (emphasis added).

against the government that is precluded by the exceptions set forth in Section 2680 of Title 28, U.S.C.[,] also is precluded against an employee . . . ."[119] Justice Marshall found these provisions clearly stated Congress' intent.[120]

Justice Stevens disagreed with the *Smith* majority's construction of section 5.[121] According to Justice Stevens, unless section 5(b)(2)(B) was intended to preserve the Gonzalez Act remedy, "it was essentially without purpose."[122] In his dissent, however, Justice Stevens never explained how a physician could violate the Gonzalez Act and, thus, be subject to a suit brought for a violation of a United States statute that authorizes action against a federal employee.[123]

In support of the plaintiffs' claims, both the Ninth Circuit and the plaintiffs advanced additional arguments, all of which Justice Marshall rejected.[124] The Ninth Circuit argued if Congress intended to extend immunity coverage under the Liability Reform Act to federal employees for tort claims arising in foreign countries, Congress would have expressly done so as it did for TVA employees in section 9 of the Act.[125] Justice Marshall concluded the Ninth Circuit misunderstood "the purpose and effect of § 9."[126] According to Justice Marshall, section 9 provides that a suit against TVA rather than a suit against the United States is the exclusive remedy for the employment-related torts of TVA employees.[127] Congress enacted section 9 simply to clarify that TVA would be substituted as the defendant in tort suits against TVA employees.[128] Thus, Justice Marshall concluded the enactment of section 9 did not support an inference either way as to Congress' intent with respect to the protection afforded federal employees by section 5.[129]

Similarly, Justice Marshall rejected the statutory argument advanced by the plaintiffs that implied repeals should be avoided.[130] The

---

119. *Id.* at 5950.

120. *Smith*, 111 S. Ct. at 1190. Disagreeing with the Ninth Circuit, Justice Marshall found no inconsistencies in the legislative history of the Liability Reform Act. According to Justice Marshall, Congress never suggested the Act "did not narrow existing *substantive* rights of recovery." *Id.* at 1185-86 n.9. The Act simply "preserved the *procedural* right to initiate an action." *Id.*

121. *Id.* at 1189-95.

122. *Id.* at 1194.

123. *Id.* at 1189. Neither Justice Marshall nor Justice Stevens was aware of what causes of action Congress sought to preserve with § 2679(b)(2)(B). *Id.*

124. *Id.* at 1188-89.

125. Smith v. Marshall, 885 F.2d 650, 655 (9th Cir. 1989).

126. *Smith*, 111 S. Ct. at 1186.

127. *Id. See supra* note 57.

128. *Smith*, 111 S. Ct. at 1186. Congress found it necessary to clarify section 9 because district courts had held that TVA has tort liability independent of the FTCA. *Id.*

129. *Id.* Comparing section 9 with section 5, Justice Marshall concluded both sections conferred the same degree of immunity to federal employees. *Id.*

130. *Id.* at 1187. *See, e.g.*, Randall v. Loftsgaarden, 478 U.S. 647, 661 (1986).

plaintiffs asserted that construing the Liability Reform Act to immunize military physicians practicing abroad results in an implied repeal of the Gonzalez Act.[131] Finding the plaintiffs' reasoning faulty, Justice Marshall explained that application of the Liability Reform Act to a tort action based on state or foreign law does not effect a repeal of the Gonzalez Act.[132] Justice Marshall noted the Gonzalez Act does not create rights for a plaintiff; the Act only serves to protect military medical personnel from malpractice liability.[133] Because Congress did not create the right of a plaintiff to sue for negligence, any subsequent limitation by Congress does not effect an implied repeal.[134]

Finally, the plaintiffs argued the Liability Reform Act only applies to federal employees not already protected from tort liability, unlike military physicians who are protected by the Gonzalez Act.[135] Justice Marshall, however, found this construction to be inconsistent with Congress' intent.[136] According to Justice Marshall, the plain language of the statute makes no distinction between employees covered by other immunity statutes and those who are not covered.[137] The Liability Reform Act clearly refers to "any employee" and contains no restricting phrases.[138] Moreover, Justice Marshall noted when Congress wanted to limit the applicability of the Act, it expressly did so.[139]

The holding in *Smith* affects more than medical malpractice cases against military physicians practicing abroad. The ruling is applicable to all claims involving an exception to the FTCA set out in section 2680.[140] At present, no remedy is available for injuries caused by the

---

131. *Smith*, 111 S. Ct. at 1187.

132. *Id.* at 1188.

133. *Id.* The Gonzalez Act was enacted partially in response to a decision in the United States Court of Appeals for the District of Columbia Circuit. That court held an Army medical officer did not have absolute immunity from civil actions arising from medical decisions. Henderson v. Bluemink, 511 F.2d 399, 403-04 (D.C. Cir. 1974). *See also* S. REP. No. 94-1264, 94th Cong., 2d Sess. 4, *reprinted in* 1976 U.S.C.C.A.N. 4443, 4445. Congress acted to meet "the serious and urgent needs of defense medical personnel by protecting them fully from any personal liability arising out of the performance of their official medical duties." *Id.* at 2, U.S.C.C.A.N. at 4444.

134. *Smith*, 111 S. Ct. at 1188.

135. *Id.* Other acts protecting government medical personnel include the Veterans Administration Medical Malpractice Act of 1965, Pub. L. No. 89-311, 79 Stat. 1154, 56-57 (codified as amended at 38 U.S.C. § 4116 (1988)); and the Public Health Service Act of 1970, Pub. L. No. 91-623, 84 Stat. 1868 (codified as amended at 42 U.S.C. § 233 (1988)).

136. *Smith*, 111 S. Ct. at 1188.

137. *Id.*

138. *Id.*

139. *Id.* at 1188-89. Congress limited the Act's scope of immunity by preserving liability for *Bivens* actions and for actions brought under a federal statute authorizing recovery against the individual employee. 28 U.S.C. § 2679(b)(2) (1990).

140. *See supra* note 16.

negligent, employment-related acts of federal employees in situations where the FTCA precludes a suit against the United States. The plaintiffs in such actions, however, are not entirely without recourse. Both the Military Claims Act[141] and Congress' system of private bills[142] may afford some relief.

The opinion in *Smith* is well reasoned and is consistent with a plain reading of the language of the Liability Reform Act. Justice Stevens could muster no support on the Court for his dissenting view.[143] Three of the five courts of appeals that previously addressed the issue agreed with the Supreme Court's reading of the Liability Reform Act. The result is a limitation on the judicial remedies available to plaintiffs in circumstances such as the Smiths. If Congress intended a contrary result, it has the authority and the responsibility to act. As Justice Stevens noted, Congress is invited "to step in and 'provide guidance.'"[144]

KATHLEEN W. CLARK

---

141. 19 U.S.C. §§ 2731-2737 (1988). By enacting the Military Claims Act, Congress authorized the military branches to pay claims up to $100,000 for, among other things, personal injury or death. The appropriate Secretary may refer a claim to the Comptroller General if he considers it meritorious and it is in excess of $100,000. *Id.* § 2733.

142. *See* Office of Personnel Management v. Richmond, 110 S. Ct. 2465 (1990). In *Richmond*, the Court explained Congress' power to handle claims was not based on any statutory authority but was founded on a belief that "the equities and circumstances of a case create a moral obligation on the part of the Government to extend relief to an individual." *Id.* at 2475 (quoting Subcommittee on Administrative Law and Governmental Relations of the House Committee on the Judiciary, 101st Cong., 1st Sess., Supplemental Rules of Procedure for Private Claims Bills 2 (1989)). The Court further stated "Congress continues to employ private legislation to provide remedies in individual cases of hardship." *Id.*

143. *Smith*, 111 S. Ct. at 1190. Concerned that injured parties would be left without a remedy, Justice Stevens stated:

> Under the Court's holding, the Liability Reform Act has closed the door to all federal and state courts for American victims of malpractice by federal health care personnel stationed abroad. No legislative purpose is achieved by that holding because these personnel are already protected from personal liability by the Gonzalez Act and the indemnity regulation. The only significant effect of this holding is to deprive an important class of potential plaintiffs of their pre-existing judicial remedy. Respondents, and other plaintiffs like them, are now precluded from pursuing their preexisting common-law claims against an allegedly negligent doctor working abroad, even though the doctor is indemnified by the Federal Government. I cannot believe that Congress intended that result.

Id. at 1195 (footnote omitted).

144. *Id.* (quoting from *Westfall*, 484 U.S. at 300).

STANDING DOCTRINE IN ANTITRUST DAMAGE SUITS, 1890 - 1975: STATUTORY EXEGESIS, INNOVATION, AND THE INFLUENCE OF DOCTRINAL HISTORY

JOHN F. HART*

* Associate Professor, Widener University School of Law. B.A., Reed College, 1977; J.D., Yale University, 1980. I wish to thank Joe Bankman, Don Buder, Alan Garfield, Gerald Neuman, and Laura Ray for helpful comments; Rosemary Crosley for word processing; Elizabeth Halkias, Lindley Harvey, Pamela Heritage, Katrina Luedtke, and Bill Rimmer for research assistance; and the librarians of the Library of Congress, Widener University, and Yale Law School for their help.

INTRODUCTION

In 1890, section 7 of the Sherman Antitrust Act authorized private treble-damage actions in broad terms: "Any person who shall be injured in his business or property by any other person or corporation by reason of anything forbidden or declared to be unlawful by this act, may sue therefore . . . ."[1] This provision has endured for a century, having been incorporated in section 4 of the Clayton Act in 1914 virtually unchanged;[2] the suits it authorizes constitute the "archetype" of the "private Attorney General" action.[3]

The effective scope of the statute's broad language for some time has been subject to the stringent restrictions of antitrust standing doctrine. Foremost among such restrictions has been the rule that only "directly" injured plaintiffs are entitled to sue for treble damages.[4] Courts and commentators working with antitrust standing doctrine have produced a conventional account of its history.[5] According to this account, the direct-injury rule constituted the courts'

---

1. Sherman Antitrust Act, ch. 647, 26 Stat. 210 (1890) (§ 7 repealed 1955).

2. *See* Clayton Act, ch. 323, 38 Stat. 731 (1914) (current version at 15 U.S.C. § 15 (1973 & Supp. 1990)).

> That any person who shall be injured in his business or property by reason of anything forbidden in the antitrust laws may sue therefor in any district court of the United States in the district in which the defendant resides or is found or has an agent, without respect to the amount in controversy, and shall recover threefold the damages by him sustained, and the cost of suit, including a reasonable attorney's fee.

Id.

3. John L. Coffee, Jr., *Rescuing the Private Attorney General: Why the Model of the Lawyer as Bounty Hunter Is Not Working*, 42 MD. L. REV. 215, 217 (1983).

4. *E.g.*, Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952). Today standing to seek treble damages is determined according to a multipart balancing test. *See* Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 535-46 (1983). Directness of injury is still a central requirement. *Id.* at 540; Blue Shield v. McCready, 457 U.S. 465, 476-78 (1982).

5. *See, e.g., Associated Gen. Contractors*, 459 U.S. at 533-34 & n.29; *Conference of Studio Unions*, 193 F.2d at 54; Randolph S. Sherman, *Antitrust Standing: From* Loeb *to* Malamud, 51 N.Y.U. L. REV. 374, 378 (1976), and decisions cited therein; Daniel Berger & Roger Bernstein, *An Analytical Framework for Antitrust Standing*, 86 YALE L.J. 809, 814 n.16, 818 n.37 (1977); Note, *Standing to Sue For Treble Damages Under Section 4 of the Clayton Act*, 64 COLUM. L. REV. 570, 581-82 (1964).

original treatment of the standing issue and dominated antitrust standing doctrine thereafter.[6] This conventional history in turn has assumed doctrinal importance, as courts have used it to explain the inconsistency between the broad language of the treble-damage provisions and the restrictions of standing doctrine.[7]

The Supreme Court's recent justification for its synthesis of standing doctrine relies on the same general historical account.[8] The Court asserts that the "federal judges who first confronted the task of giving meaning to section 7 . . . avoided a simple literal interpretation" because they understood that Congress had intended them to apply "constraints comparable to well-accepted common-law rules" that "circumscribed the availability of damages recoveries in both tort and contract litigation—doctrines such as foreseeability and proximate cause, directness of injury, certainty of damages, and privity of contract."[9]

Given the lack of explicit reference to these common-law limitations during consideration of the Sherman Antitrust Act,[10] this characterization of early antitrust standing doctrine is also crucial evidence for the Court's inference that Congress intended the courts to apply restrictive standing doctrines to section 7.[11] The enactment of section 4 of the Clayton Act in 1914, closely based on section 7 of the Sherman Antitrust Act, in turn is claimed to reflect Congress' adoption of this "judicial gloss," including the direct-injury requirement.[12] In the Court's reasoning this account of early standing doctrine legitimates not only the direct-injury requirement, but also other forms of restriction on standing elaborated by the courts in later years.[13]

This Article will assess the validity of the conventional account of the history of treble-damage standing doctrine. Close analysis of antitrust case law shows that this conventional account is considerably

---

6. *See supra* note 5.

7. *See, e.g., In re* Multidistrict Vehicle Air Pollution M.D.L. No. 31, 481 F.2d 122, 125 (9th Cir. 1973), *cert. denied*, 414 U.S. 1045 (1973); Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956).

8. *Associated Gen. Contractors*, 459 U.S. at 532-34.

9. *Id.* at 533-34. "[L]egislators familiar with these limits could hardly have intended the language of Section 7 to be taken literally." *Id.* at 533 n.28.

10. *Id.* at 533.

11. *See id.* at 530, 533-34.

12. *Id.* at 534.

13. *See Associated Gen. Contractors*, 459 U.S. at 533 n.28 ("The common law, of course, is an evolving body of law. We do not mean to intimate that the limitations on damages recoveries found in common-law actions in 1890 were intended to serve permanently as limits on Sherman Antitrust Act recoveries."). "[A]s was required in common-law damages litigation in 1890," standing issues require courts to "evaluate the plaintiff's harm, the alleged wrongdoing by the defendants, and the relationship between them." *Id.* at 535.

incorrect. The direct-injury rule was not the courts' original approach to standing issues; instead, early standing analysis generally deferred to the ordinary meaning of the broad statutory language.[14] The direct-injury rule did not dominate antitrust standing doctrine until the 1950s.[15] Indeed, the direct-injury rule conclusively displaced the ordinary-meaning statutory-construction approach only in the 1970s.[16]

Given the disparity between the conventional account and the evidence presented below, the question of how the conventional account gained authority arises as an historical problem in its own right. Secondarily, then, this Article will examine the evolution in the representation of this doctrine's history and how this historical representation influenced contemporary doctrine.

Part I will examine the period between the enactment of the Sherman Antitrust Act in 1890 and the enactment of the Clayton Act in 1914. The treble-damage case law of 1890-1914 presents a much broader range of standing issues and a much different resolution of those issues than is indicated in modern accounts. The dominant approach to standing issues was a deferential, ordinary-meaning standard of statutory construction.[17] Under this standard, denial of standing to sue could be justified only as an implementation of the statutory language or as a discrete exception to the statutory language, based on a pre-existing common-law rule that governed a category comprehending section 7 actions.[18] The general rule of standing implied by the ordinary-meaning approach to the statute was simply that any plaintiff could sue whose injury was demonstrable by nonspeculative evidence.[19] The direct-injury principle, conventionally depicted as the courts' original approach to standing issues, appeared only in ambiguous dictum in one case during this period.[20]

Part II will examine antitrust standing doctrine from 1915 to 1950. This period was one of transition. In contrast to the relative coherence and consensus of the prior ordinary-meaning case law, the case law of 1915-1950 presented variety, uncertainty, and an emphasis on practical results rather than on abstract consistency.[21] Some decisions manifested continuity with the prior case law of 1890-1914, whether in reasoning or in result; others ignored that case law altogether.[22]

---

14. *See infra* notes 68-74, 80-81 and accompanying text.
15. *See infra* notes 188-228, 266-69 and accompanying text.
16. *See infra* notes 382-83 and accompanying text.
17. *See infra* notes 68-74, 135-37 and accompanying text.
18. *See infra* notes 80-87, 135-37 and accompanying text.
19. *See infra* notes 68-78, 93-100 and accompanying text.
20. *See infra* notes 107-37 and accompanying text.
21. *See infra* notes 231-50 and accompanying text.
22. *See infra* notes 143-45, 163-67, 182-84 and accompanying text.

The most important development of this period was that the ordinary-meaning reading of the statute gradually lost its previous place as a unifying theory of antitrust standing.[23] Concrete rules of antitrust standing for certain categories of plaintiffs, originally derived from the ordinary meaning of section 7 of the Sherman Antitrust Act, were repeatedly invoked until they became the most familiar propositions of antitrust standing doctrine.[24] Because the origins of these rules had been forgotten, by the 1940s the rules were perceived as inconsistent with the ordinary-meaning approach.[25] During this period, several courts forcefully asserted the direct-injury rule but most courts ignored it.[26]

Part III will examine the period from 1951 to 1975. The direct-injury rule had languished until 1950, but activist courts abruptly revived it in the ensuing decade.[27] Although a few of their opinions asserted that the direct-injury rule always had governed antitrust standing analysis,[28] most contemporary courts clearly understood themselves to be creating new doctrine.[29] The roots of modern antitrust standing doctrine really lie in the 1950s, not in the early years of the antitrust laws.

The public-injury rule, another restrictive standing doctrine, also arose in the 1950s.[30] It seems to have shared a common impetus with the revived direct-injury rule: the lower courts' alarm at the burdens of judicial administration imposed by private antitrust litigation.[31] The public-injury rule, however, attracted less widespread support among the courts than the direct-injury rule.[32] This was probably because the public-injury rule's lack of a plausible basis in either the statute or doctrinal history revealed it more clearly as an illegitimate encroachment on the statutory remedy.[33]

The decisions of the 1950s based on the direct-injury rule were only loosely consistent, and they did not produce a stable consensus on the content of this supposed standing requirement.[34] In the 1960s many courts began a substantial reworking of the direct-injury concept: some expanded the scope of standing while others restricted it.[35] Some courts even began to base standing analysis once again on

---

23. *See infra* notes 163-74 and accompanying text.
24. *See infra* notes 143-45, 250 and accompanying text.
25. *See infra* notes 169-74 and accompanying text.
26. *See infra* notes 182-228 and accompanying text.
27. *See infra* notes 254-68 and accompanying text.
28. *See infra* note 277 and accompanying text.
29. *See infra* notes 277-88 and accompanying text.
30. *See infra* notes 289-93 and accompanying text.
31. *See infra* notes 306-15 and accompanying text.
32. *See infra* notes 296-97 and accompanying text.
33. *See infra* notes 317-28 and accompanying text.
34. *See infra* notes 355-59 and accompanying text.
35. *See infra* notes 329-54 and accompanying text.

a formal, ordinary-meaning approach to the statute.[36] Ironically, however, these courts acted without any awareness of the substantial historical basis for the ordinary-meaning approach,[37] and this approach was eventually rejected as an illegitimate innovation.[38] The end-point of this study is 1975, the eve of the Supreme Court's active innovation in this area.[39]

Part IV will address the implications of this Article's historical findings for today's antitrust standing doctrine. The legitimacy of modern antitrust standing doctrine as articulated by the Supreme Court rests fundamentally on three historical propositions: first, that in enacting the treble-damage provisions of the Sherman Antitrust Act, Congress intended the courts to develop standing restrictions comparable to the direct-injury rule;[40] second, that the courts adopted a generally applicable direct-injury rule of antitrust standing from the beginning;[41] and third, that in enacting the treble-damage provisions of the Clayton Act in 1914, Congress intended to confirm the direct-injury rule thus developed by the courts.[42] The first and third propositions run counter to the face of the statutory language and are not themselves directly supported by the legislative record;[43] both largely depend on the second proposition. The historical case for modern standing doctrine depends crucially on the proposition that the antitrust standing case law of 1890-1914 was dominated by the direct-injury rule.

This Article demonstrates that this historical proposition is incorrect. The direct-injury rule played a negligible role in antitrust case law until well after the passage of the Clayton Act;[44] the dominant approach to standing of 1890-1914 was based instead on the ordinary meaning of the statute.[45] The Supreme Court has declared that "in the absence of some articulable considerations of

---

36. *See infra* note 362 and accompanying text.

37. *See infra* notes 368-69 and accompanying text.

38. *See infra* notes 365-71, 383-84 and accompanying text.

39. Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc., 429 U.S. 477 (1977); Blue Shield v. McCready, 457 U.S. 465 (1982); Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519 (1983).

40. *Associated Gen. Contractors*, 459 U.S. at 532-33.

41. *Id.* at 532-34.

42. *Id.* at 534.

43. *See supra* note 10 and accompanying text. The only other support in the legislative history of the Sherman Antitrust Act would be "the frequent references to common-law principles," said to "imply that Congress simply assumed that antitrust damages litigation would be subject to constraints comparable to well-accepted common-law rules applied in comparable litigation." *Associated Gen. Contractors*, 459 U.S. at 533.

44. *See infra* notes 188-228, 266-69 and accompanying text.

45. *See infra* notes 61-74, 80-87 and accompanying text.

statutory policy suggesting a contrary conclusion in a particular factual setting," section 4 of the Clayton Act should be applied "in accordance with its plain language and its broad remedial and deterrent objectives."[46] Neither of the policy objectives associated with the rise of the direct-injury rule—limiting the cumulative size of damage awards assessed against particular defendants[47] and limiting the burdens imposed on the lower courts by private antitrust litigation[48]—can be regarded as inherent in section 4 of the Clayton Act. Both are flatly contrary to that section's language.[49] Because of the lack of support for a generally applicable direct-injury requirement in the statutory language,[50] the legislative history,[51] and the early case law,[52] the modern direct-injury requirement should be rejected as an "artificial limitation" on the treble-damage remedy.[53]

More generally, the revised historical picture presented in this Article should help restore to the language of the statute its former role as the fundamental source of guidance for treble-damages standing doctrine. Underlying the notion that Congress intended to give the courts wide latitude in developing standing requirements[54] is the Supreme Court's premise that for purposes of antitrust standing the broad statutory provision for treble damages "cannot mean what it says."[55] But no such argument was even proposed in the early cases, much less reflected in the early decisions.[56] The ordinary meaning of the statute was the meaning courts gave it during the period following the Sherman Antitrust Act, and which Congress tacitly ratified in the treble-damage provision of the Clayton Act.[57] Especially when the legislative record offers no support for restrictions,[58] this early construction of section 7 is powerful evidence that the "naturally broad and inclusive meaning" of the treble-damage section is entitled to virtually conclusive weight in standing analysis.[59]

---

46. Blue Shield v. McCready, 457 U.S. 465, 473 (1982).
47. *See infra* notes 302-03 and accompanying text.
48. *See infra* notes 306-15 and accompanying text.
49. *See supra* note 2.
50. *See supra* note 2.
51. *See supra* notes 10-12 and accompanying text.
52. *See supra* note 44.
53. Blue Shield v. McCready, 457 U.S. 465, 472 (1982).
54. *See supra* note 13 and accompanying text.
55. Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 531 (1983) (quoting National Soc'y of Professional Eng'rs v. United States, 435 U.S. 679, 687 (1978)).
56. *See infra* notes 68-74, 90-92 and accompanying text.
57. *See supra* notes 1-2, *infra* notes 68-74, 81-87 and accompanying text.
58. *See supra* notes 10-11 and accompanying text.
59. Blue Shield v. McCready, 457 U.S. 465, 473 (1982) (quoting Reiter v. Sonotone Corp., 442 U.S. 330, 338 (1979)).

I. 1890-1914: THE ORIGINAL UNDERSTANDING

A. Questions of Standing and the Ordinary-Meaning Principle of Statutory Construction

Antitrust standing analysis, today an essentially judicial artifact, began simply as an instance of statutory interpretation. In the years 1890-1914 courts did not differentiate between antitrust standing issues and other issues of construing the treble-damage provision, section 7 of the Sherman Antitrust Act.[60] Contemporary courts indicated that the ordinary meaning of section 7's statutory terms would normally govern.[61] Grounds for exception might arise if the pertinent terms were ambiguous,[62] or if an authoritative common-law rule applied to a category whose scope would comprehend section 7 actions, so that Congress could be considered to have intended that rule to apply.[63]

This formal reasoning was invoked in resolving a wide variety of issues arising under section 7. Because section 7 actions fell within the category of actions at law, for example, they were held to be subject to rules regarding proof of damages[64] and contractual defenses[65]

---

60. *E.g.*, Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910); United States v. American Tobacco Co., 164 F. 700 (C.C.S.D.N.Y. 1908); City of Atlanta v. Chattanooga Foundry & Pipeworks, 127 F. 23 (6th Cir. 1903), *aff'd*, 203 U.S. 390 (1906).

61. *E.g.*, Mannington v. Hocking Valley Ry., 183 F. 133, 155 (C.C.S.D. Ohio 1910) ("The first resort in all cases is to the natural, ordinary, familiar signification of the words employed. The presumption is that language has been employed with sufficient precision to disclose the intent, and, unless an examination overthrows the presumption, nothing remains but to enforce the statute as written.").

62. *Id.*; United States v. United Shoe Mach. Co., 234 F. 127, 146 (E.D. Mo. 1916) ("If the language is clear and free from ambiguity, there is nothing for the courts to construe.").

63. *E.g.*, Lowry v. Tile, Mantel & Grate Ass'n, 106 F. 38, 46 (C.C.N.D. Cal. 1900).

> It is not enough, in *an action of this kind, which is one at law*, for the plaintiffs to establish the existence of an association which comes within the inhibition of the act of congress. . . . The damages *which the law contemplates, and which the act of congress provides for*, must be reasonable damages ascertainable upon the evidence presented in the case.

Id. (emphasis added); *see* Locker v. American Tobacco Co., 218 F. 447, 448-50 (2d Cir. 1914).

64. *Lowry*, 106 F. at 46; Central Coal & Coke Co. v. Hartman, 111 F. 96, 98 (8th Cir. 1901) (the rule that damages must be proved with reasonable certainty should apply to antitrust damage actions because that rule was one of the "fundamental principles of the law of damages").

65. *See City of Atlanta*, 127 F. at 28. The courts limited the use of § 7 as a contractual defense in accordance with the "general principle of law which forbids the setting off of unliquidated damages not directly growing out of the principal

applicable to actions at law generally. Section 7 actions were litigated in federal court; therefore they were held to be governed by established federal court rules regarding service of process.[66] Because section 7 actions were statutory rather than common-law actions, pleading rules governing common-law actions were held not applicable to section 7 complaints.[67]

The courts invoked the same deferential, ordinary-meaning approach to section 7 in resolving standing issues. The (infrequent) use of the term "standing" in treble-damage opinions during this era reflects this understanding.[68] The critical issue in standing analysis, consequently, was whether the terms of section 7 pertinent to standing ("[a]ny person . . . injured . . . by reason of anything forbidden")[69] were ambiguous in this context. They were not, in the view of contemporary courts. The observation that section 7 was "very general, . . . in no wise detailing or limiting the character of injuries

---

transaction." *Id.*; *cf.* D.R. Wilder Mfg. v. Corn Prods. Ref. Co., 236 U.S. 165, 172 (1915) (Sherman Antitrust Act governed by "the elementary proposition that courts may not refuse to enforce an otherwise legal contract because of some indirect benefit to a wrongdoer"); Harriman v. Northern Sec. Co., 197 U.S. 244, 295 (1905) (complainants sought rescission of a contract previously held to have violated the Sherman Antitrust Act; Supreme Court applied "the settled rule that property delivered under an illegal contract cannot be recovered back by any party *in pari delicto*").

66. Thorburn v. Gates, 225 F. 613 (S.D.N.Y. 1915).

> Section 7 of the Sherman Antitrust Act, in providing that the defendant may be served where "found," did not intend to extend the scope of process of this court. . . . The validity of this service of process, therefore, gains nothing from the fact that the action arises under section 7 of the Sherman Antitrust Act, but is to be judged quite as though it had been an ordinary civil action . . . .

Id. at 615. *See* Frey & Son, Inc. v. Cudahy Packing Co., 228 F. 209, 210 (D. Md. 1915) ("Clearer language than that used would be required to show that Congress intended to change the rule that an officer, agent, or employe [sic] of a corporation cannot carry it into any jurisdiction in which he is not acting for it.").

67. *See* Buckeye Powder Co. v. E.I. du Pont de Nemours Powder Co., 196 F. 514, 520-21 (D.N.J. 1912); Strout v. United Shoe Mach. Co., 195 F. 313, 317 (D. Mass. 1912); Ware-Kramer Tobacco Co. v. American Tobacco Co., 178 F. 117, 120 (C.C.E.D.N.C. 1910); *cf.* Monarch Tobacco Works v. American Tobacco Co., 165 F. 774, 779 (C.C.W.D. Ky. 1908) ("like other civil actions" in Kentucky federal court, § 7 action was governed by Kentucky statutory provisions regarding specificity of pleading).

68. *E.g.*, Pidcock v. Harrington, 64 F. 821, 822 (C.C.S.D.N.Y. 1894) ("But for [section 7] no private person would have any standing in court . . . ."); *City of Atlanta*, 127 F. at 25 ("[I]f [plaintiff] has no standing to recover damages for an injury to its 'business,' it is not easy to see how it has any better standing to recover for injury to its 'property.'"); Loeb v. Eastman Kodak Co., 183 F. 704, 707 (3d Cir. 1910). It was not until the 1950s that the term "standing" was widely used to refer to the question of whether an antitrust plaintiff was eligible to bring suit. *See infra* note 286 and accompanying text.

69. Ch. 647, 26 Stat. 210 (1890).

for which a right to sue is given" did not detract from its intelligibility.[70] Rather, Courts understood section 7's broad terms to define concretely "harm . . . of the type which the statute was intended to prevent."[71] This ordinary-meaning reading of section 7 was judged to "effectuate the purposes for which it was enacted."[72] Indeed, the courts drew attention to the plain, literal aspects of its language: section 7 was "so clear and plain in its provisions that its meaning cannot be uncertain."[73] The courts' paraphrasing of section 7 tended to emphasize its broad scope.[74]

The Supreme Court's recent statement that section 7 of the Sherman Antitrust Act is like section 1 in that it "cannot mean what it says" regarding standing[75] is at odds with the early reading of

---

70. *Monarch Tobacco Works*, 165 F. at 777.

> [T]he language of section 7 is very brief and very general, prescribing no limits, except the broad one that the suit it authorizes shall be for injuries which have been suffered at the hands of any person or corporation by reason of anything forbidden or declared to be unlawful by the act.

Id. at 778.

71. Wheeler-Stenzel Co. v. National Window Glass Jobbers' Ass'n, 152 F. 864, 874 (3d Cir. 1907).

72. *Monarch Tobacco Works*, 165 F. at 778; *see, e.g.*, *City of Atlanta*, 127 F. at 27 (nonrestrictive interpretation of § 7 was identified with "the wide economic purposes of Congress"). "Congress evidently foresaw the wholesome effect of pecuniary responsibility for injuries resulting from such forbidden combinations and the courts should not devitalize the remedy by strained interpretations calculated to encourage disregard of the law." *Id.*

73. Strout v. United Shoe Mach. Co., 195 F. 313, 317 (D. Mass. 1912). Similarly, a monograph declared that § 7 was "so plain and precise in all its parts that it requires only to be attentively read in order to be understood." ALBERT H. WALKER, HISTORY OF THE SHERMAN LAW 61 (1910).

74. *E.g.*, Ware-Kramer Tobacco Co. v. American Tobacco Co., 180 F. 160, 165 (C.C.E.D.N.C. 1910) (as long as a conspiracy by defendants to restrain interstate trade was shown, "an injury 'done to the business or property' of 'any person,' by reason thereof" constituted a cause of action); *Monarch Tobacco Works*, 165 F. at 781 (§ 7's provision in "most general language" that "'any person who shall be injured in his business or property . . . by reason of anything forbidden . . . by this act' shall have a right to recover therefor" should be given effect); *Wheeler-Stenzel Co.*, 152 F. at 871 (where a defendant's combination violated Section 1 of the Sherman Antitrust Act, "an action properly accrues under the seventh section to any one who has been injured in his business or property by reason thereof"); Loder v. Jayne, 142 F. 1010, 1012 (C.C.E.D. Pa. 1906) (§ 7 authorized "*every* person injured in his business or property to bring suit against such other person or corporation, who may be engaged in any such unlawful act") (emphasis added).

75. Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 531 (1983) (quoting National Soc'y of Professional Eng'rs v. United States, 435 U.S. 679, 687-88 (1978)).

During this period the Court held that § 1's reference to "[e]very contract" could not have been intended to receive its ordinary meaning, because that would dictate an inconceivable result: the voiding of all contracts in interstate commerce. *See* Board of Trade v. United States, 246 U.S. 231, 238 (1918); Anderson v. United States, 171 U.S. 604, 616 (1898).

section 7. No early court held that the ordinary meaning of section 7's reference to "any person" would generate inconceivable results, nor did any defendant even attempt to make this argument.[76] The apparent reason for the modern view that section 7 cannot mean what it says is the premise that it would then encompass all of the "ripples of harm" that an antitrust violation would cause to "flow through the Nation's economy."[77] This objection takes section 7 out of context, however. The damages action authorized by section 7 was always subject to the "fundamental principles of the law of damages" requiring a plaintiff to prove injury by nonspeculative evidence,[78] which substantially qualified the expansiveness of the ordinary-meaning reading of section 7.[79]

1. Exceptions to the Ordinary Meaning of the Statute

Exceptions to the ordinary meaning of section 7 of the Sherman Antitrust Act were recognized if a pre-existing common-law rule had governed a discrete category that would comprehend section 7 actions, such that Congress could be considered to have intended that rule to apply.[80] Among decisions resolving standing issues under section 7, grounds for such exceptions were recognized only for two related categories of plaintiffs: stockholders and creditors bringing damage suits for losses deriving from injury to the corporation.[81]

---

76. *Compare* WALKER, *supra* note 73, at 56 (construction of § 1 threatened "trivial and vexatious suits") *with id.* at 61 (§ 7 was "so plain and precise in all its parts that it requires only to be attentively read in order to be understood").

77. Blue Shield v. McCready, 457 U.S. 465, 476-77 (1982).

78. Central Coal & Coke Co. v. Hartman, 111 F. 96, 98 (8th Cir. 1901).

> Compensation for the legal injury is the measure of recoverable damages. Actual damages only may be secured. Those that are speculative, remote, uncertain, may not form the basis of a lawful judgment. The actual damages which will sustain a judgment must be established, not by conjectures or unwarranted estimates of witnesses, but by facts from which their existence is logically and legally inferable.

Id.; *see* Locker v. American Tobacco Co., 218 F. 447, 448 (2d Cir. 1914) (it must be proved that defendants' violation of the Sherman Antitrust Act has "injured the plaintiffs and caused them damages which can be recovered in *an action at law*") (emphasis added); *id.* at 450 ("[W]e find no satisfactory proof of damages; the matter seems to be left to speculation and conjecture."); *see also* Loder v. Jayne, 142 F. 1010, 1019 (C.C.E.D. Pa. 1906) ("The items of damage claimed must be established by proof of facts from which they may be rationally inferred with reasonable certainty by the jury."), *rev'd on other grounds*, 149 F. 21 (3d Cir. 1906); Keogh v. Chicago & Northwestern Ry. 260 U.S. 156, 165 (1922).

79. *See Blue Shield*, 457 U.S. at 475 n.11 (speculativeness of plaintiff's evidence of injury classified as part of standing analysis).

80. *See supra* notes 61-63 and accompanying text.

81. Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910); Ames v. American Tel. & Tel. Co., 166 F. 820 (C.C.D. Mass. 1909); Corey v. Independent Ice Co.,

A "settled policy"[82] of corporations law—"the ordinary rule that the corporation represents the stockholder in wrongs done to the corporation"[83]—previously had barred stockholder damage suits for actionable injury reducing the value of the corporation's stock. While the language of section 7 of the Sherman Antitrust Act was "comprehensive," courts determined that Congress had not intended to overturn this prior rule.[84] Section 7 suits brought by a stockholder seeking damages for an antitrust violation reducing the value of the corporation's stock were barred because they were part of the larger category governed by the pre-existing rule; this rule of corporations law was "as applicable to a violation of the Sherman act [sic] as to any other violation of law."[85] Under the same principle of construction, some courts that denied stockholders standing to sue for treble damages indicated at the same time that stockholders would be permitted to invoke the Sherman Antitrust Act in equity actions of a type permitted by prior case law.[86] The reasons given for denying

---

207 F. 459 (D. Mass. 1913). *But see* Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 532-34 (1983). The Supreme Court has posited that in the earliest treble-damage actions the courts "avoided a simple literal interpretation" because they understood that Congress had intended them to apply "constraints comparable to well-accepted common-law rules" that "circumscribed the availability of damages recoveries in both tort and contract litigation—doctrines such as foreseeability and proximate cause, directness of injury, certainty of damages, and privity of contract." *Associated Gen. Contractors*, 459 U.S. at 532-34.

82. *Loeb*, 183 F. at 709.

83. *Ames*, 166 F. at 824; *see* JOSEPH A. JOYCE, ACTIONS BY AND AGAINST CORPORATIONS AT LAW AND IN EQUITY § 227 at 372-74 (1910).

84. *Loeb*, 183 F. at 709; *see Ames*, 166 F. at 822-23 (§ 7 of the Sherman Antitrust Act did not "by its terms affect" the question of stockholder suits, and there was "no indication of an intention of Congress to subject a defendant to independent suits by a multitude of stockholders for an act for which the statute affords redress to the corporation itself"); *see also Loeb*, 183 F. at 709 (a contrary construction would "multiply suits . . . when [the stockholders'] wrongs could have been equally well and far more economically redressed by a single suit in the name of the corporation").

85. *Ames*, 166 F. at 822; *cf.* United Copper Sec. Co. v. Amalgamated Copper Co., 244 U.S. 261, 265 (1917) ("the long settled rule under which stockholders may seek [damages on behalf of a corporation] only in a court of equity will [not] be departed from because the cause of action involved arises under the Sherman Law").

86. *See Ames*, 166 F. at 825; Corey v. Independent Ice Co., 207 F. 459, 460 (D. Mass. 1913) (given diversity, stockholders could invoke a federal court's general equity power to obtain preventive relief against a violation of the Sherman Antitrust Act harming the corporation); *see also* Mannington v. Hocking Valley Ry., 183 F. 133, 140 (C.C.S.D. Ohio 1910) (jurisdiction noted in stockholder's equity suit under Sherman Antitrust Act).

The stockholder's equitable action to protect the interests of a corporation was a familiar device in contemporary equity practice. *E.g.*, JOYCE, *supra* note 83, at § 301. Uncertainty existed as to whether a stockholder's equity bill could be maintained regarding violations of the Sherman Antitrust Act, *e.g.*, *Loeb*, 183 F. at 709, because many courts had held that the federal courts lacked jurisdiction to hear any private suits in equity arising under the Sherman Antitrust Act. *See, e.g.*, Greer, Mills &

a stockholder plaintiff standing to sue for damages to the corporation that lowered the stock's value also applied to the corporation's creditors.[87]

The rule against stockholder treble-damage suits later became recognized as authoritative.[88] Not all courts of this initial period agreed, however, that this rule of corporations law overrode the language of section 7.[89] This split of authority reflects the strong contemporary appeal of the ordinary-meaning approach to section 7.

2. Unsuccessful Objections

No standing requirements generally applicable to all treble-damage plaintiffs were established by 1914. The ordinary-meaning approach

---

Co. v. Stoller, 77 F. 1, 3 (C.C.W.D. Mo. 1896). The latter position was eventually adopted by the Supreme Court. *See* Paine Lumber Co. v. Neal, 244 U.S. 459, 471 (1917).

87. The only case during this period to present such a claim by a creditor was *Loeb*, 183 F. 704. While the standing discussion emphasizes plaintiff's status as a stockholder, it also applies to his status as a creditor. The single "main point in the case" was "whether the plaintiff can recover for the loss of his stock . . . , and of his claim as a creditor of that corporation, under Section 7 . . . ." *Id.* at 708. The conceptually cognate relation of plaintiff's status as a stockholder and his status as a creditor of the corporation is indicated at several points. *See id.* at 709 ("If a stockholder could successfully maintain such a suit, why not any creditor of the corporation? A creditor would apparently be injured in the same way and to the same extent as a stockholder, assuming that their interests were equal."). "No conspiracy or combination against [plaintiff] as a stockholder or creditor is alleged." *Id.* Before bankruptcy the right of action for damage to the corporation lowering the value of its stock "resided in the corporation" and subsequent to bankruptcy "it resided in the trustee." *See id.* This principle would bar suits by creditors as well as stockholders. *See id.*; *Ames*, 166 F. at 824 (only "the corporation or its receiver" could seek damages for injury to the corporation). *Loeb* explicitly endorsed the reasoning of *Ames*. *Loeb*, 183 F. at 710; *see also Ames*, 166 F. at 823 (court rejected plaintiff's suggestion that stockholders and corporation both be allowed to sue, because it would improperly give the shareholder priority over corporate creditors).

Contemporary treatises reflect the cognate status of stockholders and creditors with respect to this kind of claim. *Cf.* JOYCE, *supra* note 83, at §§ 301-02 (same rules govern creditors and stockholders regarding ability to bring equity suit in corporation's behalf), § 83 at 152-53 (Equity Rule 94 of federal courts applied both to stockholders and creditors).

88. *See* United Copper Sec. Co. v. Amalgamated Copper Co., 244 U.S. 261, 264-65 (1917); *see also* Corey v. Independent Ice Co., 207 F. 459, 460 (D. Mass. 1913) ("plaintiffs concede . . . that a stockholder cannot maintain a suit at law authorized by section 7 of the act for injury to the business of his corporation").

89. *See* Bigelow v. Calumet & Hecla Mining Co., 155 F. 869, 879 (C.C.W.D. Mich. 1907) ("An action at law for the recovery of damages on account of the acts sought to be enjoined would accrue to individual stockholders, under section 7 of the federal act . . . ."); Metcalf v. American School-Furniture Co., 108 F. 909, 912 (C.C.W.D.N.Y. 1901), *aff'd per curiam*, 113 F. 1020 (2d Cir. 1902); *see also* Fleitmann v. United Gas Improvement Co., 211 F. 103, 105 (2d Cir. 1914) (district court found it "unnecessary to determine whether a single shareholder can maintain an action at law under the seventh section").

to section 7 prevailed over assertions that (i) only injury to plaintiff's own business in interstate commerce was compensable;[90] (ii) the injury to plaintiff's business or property must be "an injury as recognized at common law, that is, a harm inflicted by commission of a wrong or tort;"[91] and (iii) defendant must have intended to harm plaintiff.[92]

An issue that overlapped with questions of standing was whether treble-damage plaintiffs must meet special, restrictive standards governing remoteness of causation. Causation was subject to the rule of damages, applicable to all civil litigation, that proof of damages could not rest on speculative evidence.[93] An objection to remoteness of causation going beyond this evidentiary threshold, whether the

---

90. This objection was categorically rejected in Chattanooga Foundry & Pipe Works v. City of Atlanta, 203 U.S. 390, 397 (1906) (§ 7 authorized "recovery for the damage, although the latter was suffered wholly within the boundaries of one state"). Regarding the same point the Court of Appeals had declared, "Congress evidently foresaw the wholesome effect of pecuniary responsibility for injuries resulting from such forbidden combinations and the courts should not devitalize the remedy by strained interpretations calculated to encourage disregard of the law." City of Atlanta v. Chattanooga Foundry & Pipeworks, 127 F. 23, 27 (6th Cir. 1902), *aff'd*, 203 U.S. 390 (1906). Employing the ordinary meaning of § 7 would further "the wide economic purposes of Congress." *Id.* at 27.

This holding overruled earlier decisions. *See* Dueber Watch Case Mfg. v. Howard Watch & Clock Co., 55 F. 851, 853 (C.C.S.D.N.Y. 1893); Bishop v. American Preservers' Co. 51 F. 272, 274 (C.C.N.D. Ill. 1892); *see also* Gibbs v. McNeeley, 102 F. 594, 599 (C.C.D. Wash. 1900) (under § 7 the act that injures plaintiff must be intended to affect interstate commerce).

Another of the earliest decisions held an antitrust plaintiff could not sue for damages arising out of a contract into which plaintiff voluntarily had entered. *See* Dennehy v. McNulta, 86 F. 825, 829 (7th Cir. 1898), *cert. denied*, 176 U.S. 683 (1900).

91. Wheeler-Stenzel Co. v. National Window Glass Jobbers' Ass'n, 152 F. 864, 871 (3d Cir. 1907).

92. It was not essential for plaintiff to allege "that defendant, or the dealers represented by defendant, *attempted* to deprive it of its business . . . . It is enough, that plaintiff charges that the *result* of the illegal combination was to deprive it of customers and prevent its making a profit . . . ." *Wheeler-Stenzel Co.*, 152 F. at 874 (emphasis added). Given defendant's violation of the Sherman Antitrust Act, "the plaintiff has met the requirements of the law in the declaration, by the general statement made of damage to itself by reason thereof." *Id.* at 874; *see also* Locker v. American Tobacco Co., 194 F. 232, 233 (S.D.N.Y. 1912) (cause of action under § 7 recognized where combination was "not one against the plaintiff specifically . . . but against the public generally").

In O'Halloran v. American Sea Green Slate Co., 207 F. 187 (N.D.N.Y. 1913), *rev'd on other grounds*, 229 F. 77 (2d Cir. 1915), the court stated that defendants, though not having willfully violated the act, were "presumed to have intended the necessary, natural, and known effects or consequences of their agreements and acts, and if these effects or consequences be to unduly restrain interstate trade and commerce, then the combination is illegal, and the participants are chargeable with the consequences, and are liable for the *damages resulting*." *Id.* at 189 (emphasis added).

93. *See supra* note 78 and accompanying text.

objection was based on a narrow construction of statutory terms ("injured . . . by reason of"[94]) or on a claimed Congressional intention to incorporate some restriction found in prior tort law, would present what today would be called an issue of standing. The decisions do not always reveal the precise nature of a defendant's objections, but the courts uniformly rejected restrictive concepts of causation in section 7 cases, employing the same kind of ordinary-meaning approach to the statute used for other section 7 issues.

The courts' glosses on the statute repeatedly rendered the causation element in plain, nontechnical language.[95] The threshold showing for recovery was simply "that the business or property of the plaintiff shall have been in some way injured by reason of the illegal scheme."[96] A plaintiff had to "prove that the defendant committed [an antitrust violation], and that *in consequence* he was injured in his business or property."[97] Contemporary opinions tended to keep issues of injury separate from the elaborate causal terminology used at the same time to evaluate the substantive violation, especially the interstate commerce element, even when both turned on the same evidence.[98] The only published decision from this period that asserted

---

94. Sherman Antitrust Act, Ch. 647, § 7, 26 Stat. 210 (1890).

95. *See* H.B. Marienelli, Ltd. v. United Booking Offices of America, 227 F. 165, 171 (S.D.N.Y. 1914) ("If in the execution of [defendants' conspiracy] they injure the plaintiff, the resulting damages are within the seventh section of the act."); *O'Halloran*, 207 F. at 194 (defendants' acts must have violated the statute and "the plaintiffs must have been injured in their business by such forbidden or unlawful act;" this would constitute "such a relation to [the combination's] acts as make [sic] a prima facie case of injury"); Ware-Kramer Tobacco Co. v. American Tobacco Co., 180 F. 160, 165 (C.C.E.D.N.C. 1910) (given a violation of § 1, "an injury 'done to the business or property' of 'any person,' by reason thereof constitute[d] a cause of action"), *appeal dismissed* 196 F. 1004 (4th Cir. 1912); United States Tobacco Co. v. American Tobacco Co., 163 F. 701, 712 (C.C.S.D.N.Y. 1908) (if defendants' violation "made the price excessive and unreasonable and much greater than it would have been *but for* such combination, and the plaintiff was compelled to pay that unreasonable and excessive price . . . , he was clearly injured in his property thereby") (emphasis added).

96. Monarch Tobacco Works v. American Tobacco Co., 165 F. 774, 777 (C.C.W.D. Ky. 1908).

97. Buckeye Powder Co. v. E.I. Du Pont de Nemours Powder Co., 196 F. 514, 517 (D.N.J. 1912) (emphasis added).

98. *Compare* Loewe v. Lawlor, 208 U.S. 274, 296-308 (1908) (discussion whether defendants themselves engaged in interstate trade) *with id.* at 308-09 (notes averment that by reason of those acts plaintiffs were damaged in their business and property). *Compare* H.B. Marienelli, Ltd. v. United Booking Offices of America, 227 F. 165, 168-69 (S.D.N.Y. 1914) (analysis of whether the defendant's conduct had the requisite connection with interstate commerce) *with id.* at 171 ("If in the execution of that project they injure the plaintiff, the resulting damages are within the seventh section of the act.").

Some statements that might seem to suggest qualification of the right to sue under § 7 actually refer to the connection with interstate commerce. In Ellis v.

a causational standing requirement for section 7 beyond injury-in-fact was reversed on appeal.[99] A defendant in a later case expressly

---

Inman, Poulsen & Co., 131 F. 182 (9th Cir. 1904), the court observed that defendants "possessed the power to ruin the business of any Portland contractor who imported lumber from the adjoining state, and they exercised that power . . . . [T]he restraint was the direct and necessary result of a combination made to carry out that specific purpose." *Id.* at 188-89. This discussion of whether the combination tended "directly to . . . restrain interstate commerce" went to the question of whether defendants had violated the statute, not to the relation between plaintiff and defendants. *See id.* at 186.

In Minnesota v. Northern Sec. Co., 194 U.S. 48 (1904), the Court held the state lacked standing to bring an action in equity under the Sherman Antitrust Act. The state alleged it would be injured in its "proprietary interests" (as an owner of land, and as a taxing authority) by a consolidation of two principal railroads in the state. *Id.* at 68-70. The Court referred in passing to the treble-damage remedy of § 7:

> The injury on account of which the present suit was brought is at most only remote and indirect; such an injury as would come alike, although in different degrees, to every individual owner of property in a State by reason of the suppression . . . of free competition between interstate carriers . . . ; not such a direct, actual injury as that provided for in the seventh section . . . .

Id. at 70. In context, the reference to directness of injury almost certainly refers to the causal sufficiency of the state's claim of future injury. The chain of causation alleged would lead from the expected reduction in competition between railroads to the level of land prices and, further, to "the general prosperity and business success of its citizens," *id.* at 70; this was an unusually complex theory of injury, which would depend on speculative proof. *See infra* notes 77-78 and accompanying text. Moreover, this passage seems not to have been understood as referring to standing: it was not cited by later standing decisions, even those using a direct-injury analysis.

99. *See* Thomson [sic] v. Union Castle Mail S.S. Co., 149 F. 933, 935-36 (C.C.S.D.N.Y. 1907), *rev'd*, Thomsen v. Union Castle Mail S.S. Co., 166 F. 251 (2d Cir. 1908), *aff'd sub nom.* Thomsen v. Cayser, 243 U.S. 66 (1917). The district court, in dismissing the action, held that "no private person can recover damages against the members of the combination except such as *naturally flow from and are proximately caused by* the action of the combination." *Id.* (emphasis added). Plaintiffs were shippers of freight; the defendants were a combination of shipping lines carrying freight between New York City and South Africa. *Id.* at 933. The combination fixed uniform rates and effectively excluded competing carriers by offering substantial rebates, forfeitable if the shipper shipped freight on other lines or even if their consignees received freight shipped on other lines. *Id.* at 933-34. Plaintiffs lost their entitlement to such rebates, apparently because one of their consignees received freight from a competing ship, and sued therefor under the Sherman Antitrust Act. *See id.* at 934. That the district court's remarks concerned standing is shown by its statement that if "this case had been promoted by the United States, or even by a shipowner who, by the combined action of the defendants, had been prevented from freely engaging in commerce . . . very different questions would have been presented . . . ." *Id.* at 934.

The Second Circuit not only reversed the dismissal, but also implicitly rejected the terms in which the issue had been posed. Thomsen v. Union Castle Mail S.S. Co., 166 F. 251 (2d Cir. 1908), *aff'd sub nom.* Thomsen v. Cayser, 243 U.S. 66 (1917). Given the defendants' violation of the Sherman Antitrust Act, the remaining question under § 7 was simply "whether the plaintiffs were *thereby injured* in their business or property." *Id.* at 253 (emphasis added).

conceded that proximate cause was not essential in a section 7 action.[100]

For some courts, the inclination to read section 7 according to its ordinary meaning rather than adopt standing restrictions may have been further reinforced by the common-law principle that private persons suffering harm as a result of a statutory violation could recover damages from the violator, even where no statute provided a remedy.[101] Several section 7 decisions refer to this principle.[102] Although at common law an injury resulting from a combination in restraint of trade would not have been actionable, this was only because "such agreements were not criminal or unlawful by the common law."[103] Section 1 of the Sherman Antitrust Act had made

---

100. Ware-Kramer Tobacco Co. v. American Tobacco Co., 180 F. 160, 165 (C.C.E.D.N.C. 1910) (defendant's brief stated that a violator of the Sherman Antitrust Act will be liable only if "the unlawful conduct is the proximate cause of, *or results in*" harm to plaintiff) (emphasis added), *appeal dismissed*, 196 F. 1004 (4th Cir. 1912).

101. That is, the common law would have recognized such an action (for single damages) on the basis of a violation of §§ 1 or 2 of the Sherman Antitrust Act, even if the Sherman Antitrust Act had not explicitly provided a private remedy. *Cf.* United States v. Addyston Pipe & Steel Co., 85 F. 271, 279 (6th Cir. 1898) (Taft, J.) ("Contracts that were in unreasonable restraint of trade at common law were not unlawful in the sense of being criminal, or *giving rise to a civil action for damages in favor of one prejudicially affected thereby*, but were simply void . . . .") (emphasis added); City of Atlanta v. Chattanooga Foundry & Pipe Co., 101 F. 900, 909 (C.C.E.D. Tenn. 1900) ("The commission of an act specifically forbidden by law . . . is generally equivalent to an act done with intent to cause wrongful injury.") (quoting SIR FREDERICK POLLOCK, TORTS 23 (5th ed.)), *rev'd*, 127 F. 23 (6th Cir. 1902), *aff'd*, 203 U.S. 390 (1906).

102. Metcalf v. American School-Furniture Co., 108 F. 909, 912 (C.C.W.D.N.Y. 1901) (§ 7 was "declaratory of a common-law right which existed in favor of parties injured by wrongs enumerated in other sections of that act, and confers jurisdiction to seek a remedy, and with treble damages, in a federal tribunal"); Wheeler-Stenzel Co. v. National Window Glass Jobbers Ass'n, 152 F. 864, 873 (3d Cir. 1907) ("[I]f such combinations were illegal, in the strict sense of that word, and constituted a public wrong, whether by common law or by statute, one who incurred substantial loss thereby suffered a legal injury for which a private action would lie. . . . In a common-law jurisdiction, express statutory provision of a right of action for damage resulting from a violation of law, would not have been necessary."); *Ware-Kramer Tobacco Co.*, 180 F. at 165.

The stated principle implies that actions for violation of the Sherman Antitrust Act could be brought in state court, although § 7 conferred subject matter jurisdiction only on federal courts. *See* Blindell v. Hagan, 54 F. 40, 41 (C.C.E.D. La.) (court enjoined a Sherman Antitrust Act violation because, given diversity jurisdiction, "the complainants may urge before this court any grievance which they may have in law or equity as fully as they could do in the courts of a state"), *aff'd*, 56 F. 696 (5th Cir. 1893); *cf.* Straus v. American Publishers' Ass'n, 231 U.S. 222, 237 (1913) (reserved the question "whether an original action can be maintained in the state courts seeking an injunction and to recover damages under the Sherman Law").

103. *Wheeler-Stenzel Co.*, 152 F. at 871-73 (3d Cir. 1907) (citing Mogul S.S. Co. v. McGregor, 1892 L.R. App. Cas. 25). *Mogul Steamship* implied that "if such

restraints of trade "both unlawful and criminal,"[104] however, and it was "manifestly the correct understanding of the seventh section . . ., that where a man is harmed in his business or property by a violation of the act, he has suffered a legal injury and is entitled to his action therefor."[105]

B. Evidence of a Direct-Injury Standing Requirement

The foregoing discussion demonstrates that the dominant approach to standing issues among the early cases was based on an ordinary-meaning approach to section 7. Because the conventional historical representation has equated the standing analysis of this period with a general direct-injury rule,[106] however, the evidence presented in favor of this conventional view will now be considered. Only four decisions from 1890-1914[107] have been cited as examples of a generalized direct-injury rule.[108] In all four, plaintiffs whose losses derived from their relation to an injured corporation—as stockholders, creditors, employees, officers, or directors—were denied standing to sue for their losses under section 7. Each of these decisions, however, confirms the authority of the ordinary-meaning standard; only one of them contains even ambiguous dictum supporting a direct-injury rule.

(i) *Ames v. American Telephone & Telegraph Co.*[109]

Plaintiff was a stockholder seeking treble damages for a reduction in value of his stock, deriving from injury to the corporation.[110] The

---

combinations were illegal, in the strict sense of that word, and constituted a public wrong, whether by common law or by statute, one who incurred substantial loss thereby suffered a legal injury for which a private action would lie." *Id.* at 873. A similar reading of *Mogul Steamship* is found in United States v. Addyston Pipe & Steel Co., 85 F. 271, 286 (6th Cir. 1898), *aff'd*, 175 U.S. 211 (1899).

The *Mogul Steamship* case was extensively discussed in the early antitrust cases. *See, e.g.*, Greer, Mills & Co. v. Stoller, 77 F. 1, 7-8 (C.C.W.D. Mo. 1896); National Fireproofing Co. v. Mason Builders' Ass'n, 169 F. 259, 265 (2d Cir. 1909).

104. *Wheeler-Stenzel Co.*, 152 F. at 873.

105. *Id.* 152 F. at 874 (citing Chattanooga Foundry and Pipe Works v. City of Atlanta, 203 U.S. 390 (1906)); *see id.* at 873 (it was "obvious and unavoidable, that the private right of action . . . must be *commensurate with the extent of the illegality* thus by [the Sherman Antitrust Act] established") (emphasis added).

106. *See supra* notes 5-7 and accompanying text.

107. *See* Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910); Ames v. American Tel. & Tel. Co., 166 F. 820 (C.C.D. Mass. 1909); Corey v. Boston Ice Co., 207 F. 465 (D. Mass. 1913); Corey v. Independent Ice Co., 207 F. 459 (D. Mass. 1913).

108. *See, e.g.*, Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 533-34 & n.29 (1983); Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 n.1 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).

109. 166 F. 820 (C.C.D. Mass. 1909).

110. *Id.* at 821-22.

court explained denial of standing as a discrete exception to an ordinary-meaning reading of section 7, based on the conclusion that Congress had not intended to alter a pre-existing rule governing stockholder damage suits.[111] The court made no claims regarding congressional intention to exclude any other kind of plaintiff, nor did the court refer to other kinds of plaintiffs.[112] No such exclusions were even implied, given the distinctive characteristics of the stockholder-corporation relationship pointed out in the opinion.[113]

(ii) *Loeb v. Eastman Kodak Co.*[114]

Plaintiff was a shareholder, creditor, and employee of a company driven out of business by defendant; he sought treble damages for the loss in value of his stock shares and for his unsecured claims against the company.[115] Most of the court's reasoning is plainly recognizable as an example of the ordinary-meaning standard of construction.[116] Although the language of section 7 was "comprehensive," the "existing law" of corporations had barred stockholder damage suits for harm to the corporation lowering the value of stock, and the question posed by the court was whether Congress intended section 7 to override this "settled policy of the law."[117]

The same passage contains a reference to directness of injury[118] that has been read as asserting an independent, generally applicable principle of antitrust standing: that only plaintiffs whose injury is "direct" have standing.[119] While the sentences in question do not exclude such a reading, the context offers substantial reasons to

---

111. *Id.* at 822-23. *See supra* notes 82-86 and accompanying text.

112. *See Ames*, 166 F. at 822-23.

113. Because the plaintiff stockholder's losses duplicated those of the corporation, a single suit by the corporation would make plaintiff and all other stockholders whole. *See id.* at 822-24. Individual recovery by plaintiff would potentially jeopardize the rights of creditors and other stockholders, as well as expose the defendant to duplicative recovery. *Id.*

114. 183 F. 704 (3d Cir. 1910).

115. *Id.* at 707.

116. *Id.* at 709.

117. *Id.*

118. *Id.*

> Moreover, it is manifest that the plaintiff did not receive any direct injury from the alleged illegal acts of the defendant. No conspiracy or combination against him as a stockholder or creditor is alleged. The injury complained of was directed at the corporation, and not the individual stockholder. Hence any injury which he, as a stockholder, received was indirect, remote, and consequential.

Loeb, 183 F. at 709.

119. *E.g.*, Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952); Randolph S. Sherman, *Antitrust Standing: From* Loeb *to* Malamud, 51 N.Y.U. L. REV. 374, 379 (1976).

conclude that the court was merely addressing the standing of stockholders of the type presented by this plaintiff. The term "direct" commonly appeared as a term of corporate case law,[120] and its use here is fully explainable as such. The plaintiff suffered no "direct injury" from the defendants, because the injury was "directed at the corporation."[121] This is why, under principles of corporations law existing before the Sherman Antitrust Act, such a stockholder "would have been without direct relief."[122] Some courts recognized an exception to the rule of corporations law barring stockholder damage suits where defendants had intended to inflict harm on plaintiffs, even though the corporation was an intervening victim.[123] In *Loeb*, the court's point was that no variants of the general rule applied to this plaintiff, so the type of harm suffered by this stockholder was plainly not one for which Congress would have intended to provide a remedy.[124]

Other aspects of the opinion reinforce this reading. The court never explicitly addressed the general issue of antitrust standing principles, and its main objection to antitrust standing—multiple separate suits by plaintiffs whose injuries "could have been equally well and far more economically readdressed by a single suit in the name of the corporation"[125]—would not apply to plaintiffs other than stockholders and creditors.[126] Any reference to antitrust plaintiffs

---

120. *E.g.*, 6 SEYMOUR N. THOMPSON & JOSEPH W. THOMPSON, COMMENTARIES ON THE LAW OF CORPORATIONS § 4560 at 446, § 4562 at 450-51 (3d ed. 1927).

121. *Loeb*, 183 F. at 709.

122. *Id.*

123. *E.g.*, United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574 (2d Cir. 1916); *see* Note, *Standing to Sue for Treble Damages Under Section 4 of the Clayton Act*, *supra* note 5, at 582 (the plaintiff in *Loeb* "had not brought himself within the direct action exception to the rule permitting only derivative suits," for he had not "claimed a conspiracy . . . against him as a stockholder or creditor"); *cf.* Green v. Victor Talking Mach. Co., 24 F.2d 378, 381 (2d Cir. 1928) (no exception to general rule recognized); Ames v. American Tel. & Tel. Co., 166 F. 820, 824 (C.C.D. Mass. 1909) ("It is not necessary . . . to speculate as to the possibilities and as to the conceivable right of stockholders to maintain individual actions.").

124. *See Loeb*, 183 F. at 709.

125. *Id.* at 709.

126. *Id.* A stockholder typically would be injured because harm to the corporation lowered the value of his stock. A creditor would be injured because harm to the corporation prevented the corporation from paying its debt. In both cases the individual's injury would essentially duplicate that of the corporation, and a successful damages suit by the corporation would make these derivatively injured stockholders and creditors whole. *See Ames*, 166 F. at 822-23. Other kinds of indirectly injured parties would not typically stand in such a relation to more directly injured parties through whom the harm was sustained. *See* Congress Bldg. Corp. v. Loew's Inc., 246 F.2d 587, 594 (7th Cir. 1957).

In two antitrust treatises published shortly afterwards, *Loeb* was cited only as bearing on suits by stockholders or creditors; neither treatise mentioned a general direct-injury rule. *See* JOSEPH A. JOYCE, A TREATISE ON MONOPOLIES AND UNLAWFUL

beyond the corporate context would be a striking *non sequitur*, unnecessary to the holding, and would render much of the preceding reasoning superfluous. Moreover, if *Loeb*'s rule were meant to apply to plaintiffs other than stockholders and creditors of corporations, we would expect the court to invoke some generally applicable rule of law from before the Sherman Antitrust Act and assert that Congress had intended that rule to apply. The only authority cited by *Loeb* for its holding addressed nothing outside the stockholder context,[127] and *Loeb* explicitly endorsed that decision's reasoning.[128]

A general direct-injury principle of antitrust standing would have been a novelty at this time and contrary to existing precedents.[129] Yet no sign exists that the court considered itself to be breaking new ground in *Loeb*. If the court were stating a general rule of antitrust standing, one might expect the opinion to cite pertinent case law and to address the sharply contrary language in the Third Circuit's most recent section 7 opinion, especially because one judge served on both panels.[130]

Finally, *Loeb*'s remarks about directness are both preceded and followed by discussion of the pertinent law of stockholder suits as it existed before the Sherman Antitrust Act.[131] In light of the surrounding discussion, a terse digression away from the main point to make an unprecedented and unsupported pronouncement regarding non-stockholder plaintiffs seems unlikely. The entire passage is more naturally read as treating only the corporate context before and after the Sherman Antitrust Act.

(iii) *Corey v. Independent Ice Co.*[132]

In this case the court dismissed the treble-damages suit of several stockholders whose stock had lost value because of an injury defend-

---

COMBINATIONS OR RESTRAINTS §§ 160, 162 (1911); W. W. THORNTON, A TREATISE ON THE SHERMAN ANTI-TRUST ACT §§ 413, 466 [sic] (should be 446) (1913).

127. The only antitrust opinion cited in *Loeb* is Ames v. American Telephone & Telegraph Co., 166 F. 820 (C.C.D. Mass. 1909). In *Ames*, a stockholder suing for damages resulting from injury to the corporation was denied standing based on the pre-existing rule of corporation law barring such suits. *See supra* notes 109-13 and accompanying text.

128. *Loeb*, 183 F. at 710.

129. *See supra* notes 68-74, 80-81, 90-91 and accompanying text.

130. Wheeler-Stenzel Co. v. National Window Glass Jobbers' Ass'n, 152 F. 864 (3d Cir. 1907). In *Wheeler Stenzel* the court specifically had rejected the proposition that defendants must have directed their conspiracy at plaintiff to support recovery. *Id.* at 873-74. It was not "necessary, as urged by defendant, that plaintiff should allege that defendant . . . attempted to deprive it of its business . . . . It is enough, that plaintiff charges that the *result* of the illegal combination was to deprive it of customers and prevent its making a profit . . . ." *Id.* at 874 (emphasis added). Given a combination violating the Sherman Antitrust Act, the plaintiff's "general statement made of damage to itself by reason thereof" sufficiently alleged injury. *Id.* Judge Buffington sat on both the *Wheeler-Stenzel* and *Loeb* panels.

131. *Loeb*, 183 F. at 709.

132. 207 F. 459 (D. Mass. 1913).

ants inflicted upon the corporation. The decision cites *Loeb v. Eastman Kodak Co.*,[133] but only for the rule that a stockholder cannot bring "a suit at law authorized by section 7 . . . for injury to the business of his corporation whereby the value of his stock is impaired."[134] The court does not refer to directness in its reasoning and states no standing principles applicable to plaintiffs other than stockholders.

(iv) *Corey v. Boston Ice Co.*[135]

Two plaintiffs filed suit in their capacity as directors and officers of the injured corporation.[136] The court's dismissal of these complaints was presented not as an exception to the language of section 7, but rather as an application of the ordinary-meaning standard, based on plaintiffs' lack of pertinent "business or property."[137] The opinion contains no reference to any principle of directness.

II. 1915-1950: EVOLUTION

The Clayton Act of 1914 left the statutory basis of standing doctrine unchanged;[138] the substantial evolution of antitrust standing doctrine in later decades was without impetus from Congress. Section 4 established a private treble-damages remedy in substantially the same terms as section 7 of the Sherman Antitrust Act,[139] comprehending injury to the business or property of "any person . . . by reason of anything forbidden in the antitrust laws."[140] (Although section 7 of the Sherman Antitrust Act would seem to have been

---

133. 183 F. 704 (3d Cir. 1910).

134. *Independent Ice Co.*, 207 F. at 460.

135. Corey v. Boston Ice Co., 207 F. 465 (D. Mass. 1913).

136. *Id.* at 466.

137. *Id.* at 466.

> [N]o injury to any business or property of either plaintiff appears from any of the facts alleged . . . . However long they had held their respective offices, or however frequently they may have been re-elected, there is nothing to show that they had any right to expect that they would be chosen again at this or any given election, nor to show any right of *property* in the offices mentioned or the salaries attached thereto, or any such interest in them after the dates of the meetings in 1908 as entitled them to say that failure to elect them to those offices was an injury to their *business* within the meaning of section 7 . . . .

Id. (emphasis added).

138. Clayton Act, ch. 323, 38 Stat. 731 (1914) (codified as amended at 15 U.S.C. § 15 (1973 & Supp. 1990)).

139. Sherman Antitrust Act, ch. 647, § 7, 26 Stat. 210 (1890) (repealed 1955); *see supra* text accompanying notes 1-2.

140. 38 Stat. 731 (1914) (current version at 15 U.S.C. § 15). *See supra* note 2.

superseded, it long continued to be invoked separately as a basis for treble-damage actions based on violation of the Sherman Antitrust Act.[141])

The dominant approach to standing questions during the period 1890-1914 rested on statutory construction principles. Courts viewed the language of section 7 as concretely meaningful and read this section according to its ordinary meaning in the absence of some discrete exception justified by implied congressional intent.[142] The years from 1915 to 1950 present a transition between the ordinary-meaning convention of 1890-1914 and the direct-injury convention of the 1950s. In contrast to the relative coherence and consensus that characterized these preceding and following periods, the standing case law of 1915-1950 presents doctrinal variety and vagueness. Some decisions manifest continuity with the prior case law of 1890-1914, whether in reasoning or in result; others ignore it altogether.

A. Reflections of Continuity with the Ordinary-Meaning Convention

The decisions of 1915-1950 that reflect the reasoning or the results of prior case law generally fall within the main tradition in standing case law. These are the only decisions of 1915-1950 that maintain a connection to the prior consensus. They are the most prominent in contemporary citations, and they also predominate among the decisions that are cited by later courts.

1. The Rise of Standing Rules for Categories of Plaintiffs

Perhaps the most prominent reflection of the early case law arose in decisions that formulate particularistic rules of standing, based on prior decisions rejecting certain categories of plaintiffs. The newer cases, however, fail to reiterate or refer to the reasoning of those earlier decisions. For example, the rule barring stockholders from seeking treble damages for losses from injury to their corporation originally had been justified as an exception to the ordinary meaning of section 7 of the Sherman Antitrust Act, based on the acknowledged authority of the same rule in corporations law.[143] But subsequent decisions on this point simply stated the rule based on the authority of prior decisions (not necessarily the original ones) without reiter-

---

141. *E.g.*, Georgia v. Evans, 316 U.S. 159 (1942). Section 7 of the Sherman Antitrust Act was repealed in 1955. Ch. 283, § 3, 69 Stat. 283 (1955).

142. *See supra* notes 68-74, 80-87 and accompanying text.

143. Ames v. American Tel. & Tel. Co., 166 F. 820, 822-23 (C.C.D. Mass. 1909); *see* Loeb v. Eastman Kodak Co., 183 F. 704, 709 (3d Cir. 1910).

ating the original underlying ordinary-meaning reasoning.[144] Courts no longer presented the rule about stockholders as an exception to otherwise governing statutory language, based on a determination that Congress had intended to incorporate a particular common-law rule. Instead, this rule was presented as separate and free-standing, with an unspecified relation to the pertinent statutory language.

Similarly, the courts formulated rules regarding creditors, officers, and directors based on prior decisions that had resolved the issue by application of statutory construction principles.[145] For each of these categories of plaintiffs, as for stockholders, citations to the earlier cases replaced the original statutory exegesis.[146] This line of cases effectively represented standing case law as a collection of concrete rules barring certain categories of plaintiffs, not as a doctrine containing a unifying general principle subject to discrete exceptions. For convenience, the rules of standing barring treble-damage suits by stockholders, creditors, officers, and directors for injury deriving from harm to the corporation will be referred to below as "category rules." The category rules were at first substantively neutral, but through repetition they acquired a life of their own. They became established enough to support exclusion of factually comparable plaintiffs by analogy on the facts, without justifying the results by either statutory language or inferred congressional intent.[147]

The substantive impact of representing standing doctrine as a collection of concrete rules was manifested in *Westmoreland Asbestos Co. v. Johns-Manville Corp.*[148] One plaintiff was injured as the landlord of another victim.[149] The court's reasoning, instead of starting exegetically from the statutory language, drew a factual analogy between this claim and those claims falling under the estab-

---

144. *E.g.*, Roseland v. Phister Mfg., 125 F.2d 417, 419-20 (7th Cir. 1942); Westmoreland Asbestos Co. v. Johns-Manville Corp., 30 F. Supp. 389, 391 (S.D.N.Y. 1939), *aff'd per curiam*, 113 F.2d 114 (2d Cir. 1940); Gerli v. Silk Ass'n of Am., 36 F.2d 959, 960 (S.D.N.Y. 1929); Corey v. Independent Ice Co., 207 F. 459, 460 (D. Mass. 1913).

145. The rule barring treble-damage suits by creditors of injured corporations originated in *Loeb*, 183 F. 704. *See supra* note 87 and accompanying text. The rule barring suit by the officers and directors of injured corporations originated in Corey v. Boston Ice Co., 207 F. at 466. *See supra* notes 135-37 and accompanying text. Both holdings were later stated simply as rules, with no accompanying reasoning. *See, e.g., Gerli*, 36 F.2d at 960; *Roseland*, 125 F.2d at 419-20; *Westmoreland Asbestos Co.*, 30 F. Supp. at 391; Annotation, *Who May be Regarded as Injured in His Business or Property Within Provisions of Antitrust Acts as to Person Who May Recover Damages Resulting from Violation of the Acts*, 139 A.L.R. 1017-18 (1942).

146. *See supra* notes 126, 145 and accompanying text.

147. *See supra* note 145 and accompanying text.

148. 30 F. Supp. 389 (S.D.N.Y. 1939), *aff'd per curiam*, 113 F.2d 114 (2d Cir. 1940).

149. *Id.* at 390-91.

lished category rules from the corporate context as a basis for dismissing the landlord's claim.[150] Another, subtler innovation based on factual analogy was *Westmoreland*'s restatement of the old rule barring stockholder suits for loss in value of stock.[151] The court stated that this rule covered all treble-damage claims by stockholders, barring even claims for harm that did not duplicate the corporation's harm[152] and claims where the harm to the corporation was intended to harm the plaintiff stockholder.[153]

Perhaps another consequence of treating the category rules as judicial rather than statutory artifacts, also illustrated in *Westmoreland*, was to preserve the officer rule from reconsideration.[154] The original reasoning had been that corporate officers lacked "business or property" in their employment.[155] In 1926, however, the Supreme Court held that the loss of employment by workers constituted injury to business or property.[156] If the original, ordinary-meaning reasoning of the officer rule had been reconsidered in light of this holding, it likely would have been rejected.[157] Instead, grouping the concrete rules together lent reciprocal validation and distracted attention from

---

150. *Id.* at 391. The court did not purport to give a rule for all landlords; it decided this objection by observing that this landlord's claims were "more remote" than those held to be unrecoverable in prior cases. *Id.*

151. *See id.*; *see also supra* notes 82-85, 121-23 and accompanying text.

152. Westmoreland Asbestos Co. v. Johns-Manville Corp., 30 F. Supp. 389, 391 (S.D.N.Y. 1939), *aff'd per curiam*, 113 F.2d 114 (2d Cir. 1940). While the broader holding in *Westmoreland* would have seemed unexceptional as an analogy on the facts, it was outside the scope of the original rule and its rationale: the corporation did not represent the stockholder as to losses that did not duplicate the corporation's harm.

153. *See id.* An allegation that defendants had harmed the corporation to injure the controlling stockholders in their entrepreneurial function was held to support a cause of action in United Copper Securities Co. v. Amalgamated Copper Co., 232 F. 574 (2d Cir. 1916). Plaintiffs in that case were stockholders, but they were also engaged in "organizing, promoting, and financing companies for mining, dealing in, and shipping copper." *Id.* at 577. The requirement of direct injury was met because defendants, in acquiring (and thus harming) companies in which plaintiffs' assignors owned stock, had intended to injure them. *Id.* at 577-78. In *Westmoreland*, similarly, the acts of defendants were allegedly intended to injure all of the plaintiffs. *Westmoreland Asbestos Co.*, 30 F. Supp. at 390.

154. *See Westmoreland Asbestos Co.*, 30 F. Supp. at 391.

155. *See supra* notes 135-37 and accompanying text.

156. Anderson v. Shipowners Ass'n, 272 U.S. 359, 363 (1926).

157. It could not be argued plausibly that the narrow definition of property in Corey v. Boston Ice Co., 207 F. 465, 466 (D. Mass. 1913), represented a consensus contemporary with the Sherman Antitrust Act. *See* Allen v. Commonwealth, 74 N.E. 287, 288 (Mass. 1905), *quoted in* Roseland v. Phister Mfg., 125 F.2d 417, 419 (7th Cir. 1942). As of the 1940s, the only other holdings for the rule against officer suits were found in Gerli v. Silk Ass'n of Am., 36 F.2d 959, 960-61 (S.D.N.Y. 1929), and *Westmoreland Asbestos Co.*, 30 F. Supp. at 391. Yet *Westmoreland*'s holding rested solely on *Gerli*, *Westmoreland Asbestos Co.*, 30 F. Supp. at 391, and *Gerli*'s holding rested solely on *Boston Ice Co. Gerli*, 36 F.2d at 960-61.

the origins of any particular rule.[158] Although the officer rule had originated as a construction of the statutory terms rather than an implied exception to them like the stockholder and creditor rules,[159] once the original exegesis had been elided the three were commonly grouped together when courts referred to them.[160]

In turn, the survival of the officer rule greatly extended the range of analogies to be drawn from the long-established categories of exclusion. The commonplace reference to these three rules as a group would have suggested that the basis for analogy must be a feature that stockholders, creditors, and officers all had in common (which would support a broad analogy)[161] rather than merely a feature that stockholders and creditors had in common (which would suggest a narrow analogy).[162]

2. The Uncertain Status of Ordinary-Meaning Reasoning

The second form of continuity with the prior ordinary-meaning convention was the use of an ordinary-meaning gloss on the statutory treble-damage language in addressing standing objections.[163] The ordinary-meaning reasoning survived, however, only in a diminished form. It no longer served, even for its proponents, as a unifying doctrinal principle. By the 1940s, the relation between this reasoning and the category-based rules evidently had become unclear.

The most discursive illustration of this uncertainty is presented by *Roseland v. Phister Manufacturing*.[164] The court noted that the treble-damage language was "general and all-inclusive" and decided that "the word business was used in its ordinary sense and with its

---

158. *See Roseland*, 125 F.2d at 419.

159. *See Gerli*, 36 F.2d at 960.

160. *E.g.*, *Roseland*, 125 F.2d at 419; *Gerli*, 36 F.2d at 960.

161. *See, e.g.*, East Orange Amusement Co. v. Vitagraph Inc., Trade Reg. Rep. (CCH) ¶ 52,965 (D.N.J. June 24, 1943). The court found the injury to plaintiff was not "derivative" like those of "stockholders, employees, creditors and corporate officers." *Id.* at 53,566.

162. *See supra* notes 113, 126 and accompanying text.

163. Mandeville Island Farms, Inc. v. American Crystal Sugar Co., 334 U.S. 219, 236 (1948) ("The Act is comprehensive in its terms and coverage, protecting all who are made victims of the forbidden practices by whomever they may be perpetrated."); United States v. Cooper Corp., 312 U.S. 600, 605 (1941) (language of § 7 of the Sherman Antitrust Act should be read "in its ordinary and natural sense"); *Roseland*, 125 F.2d at 419-20; Klein v. Sales Builders, Inc., 1950-1951 Trade Cas. (CCH) ¶ 62,600 (N.D. Ill. Apr. 3, 1950); Camrel Co. v. Paramount Film Distrib. Corp., 1944-1945 Trade Cas. (CCH) ¶ 57,233 (S.D.N.Y. Mar. 30, 1944); Kentucky-Tennessee Light & Power Co. v. Nashville Coal Co., 37 F. Supp. 728, 735-36 (W.D. Ky. 1941), *aff'd sub nom.* Fitch v. Kentucky-Tennessee Light & Power Co., 136 F.2d 12 (6th Cir. 1943).

164. 125 F.2d 417 (7th Cir. 1942).

usual connotations."[165] Defining "business" as "the employment or occupation in which a person is engaged to procure a living,"[166] and "accepting the words of Congress at their face value,"[167] the court held that plaintiff's position as a sales agent constituted business, and that plaintiff could seek damages for harm to it.[168]

Other aspects of the *Roseland* opinion undercut this seeming vindication of the ordinary-meaning premise, however. The court noted that stockholders, creditors, officers, and directors were not eligible to seek treble damages but offered no explanation for this proposition other than citation to case law.[169] The court's holding suggested a troublesome question: how the statutory language could at the same time provide a right of action to a sales manager and deny it to a corporate officer.[170] The court failed to explain this anomaly, except by a somewhat evasive appeal to factual differences.[171]

This implicit tension on the level of holdings in *Roseland* was matched on the level of principle. By saying so much in favor of an ordinary-meaning gloss and then failing to address the place of the categorical rules in the statutory scheme, *Roseland* implied not only that the officer rule was authoritative,[172] but also that the category rules could not be reconciled with the ordinary-meaning premise. In doing so, *Roseland* impliedly confirmed that the category rules legitimately rested on some basis other than the statute.[173] The category rules had now achieved an authority independent of their origins. They were now clearly more than shorthand representations of the original ordinary-meaning reasoning, and they were not even associated with such reasoning.

---

165. *Id.* at 419.

166. *Id.* (quoting Allen v. Commonwealth, 74 N.E. 287, 288 (Mass. 1905)). The *Roseland* court also defined "business" as "persistent human efforts which have for their end pecuniary reward." *Roseland*, 125 F.2d at 419.

167. *Roseland*, 125 F.2d at 419.

168. *Id.* at 419-20.

169. *Id.* at 419.

170. *See* Comment, *Third-Party Recovery For Injury To Economic Interests—A Common-Law Problem In Interpreting the Antitrust Laws*, 21 U. CHI. L. REV. 709, 711 (1954) ("The cases . . . do not make clear why the interest of a sales agent is considered to be less remote than those of the other third-party plaintiffs . . . .").

171. *Roseland*, 125 F.2d at 419 ("plaintiff does not seek to recover as a creditor or stockholder"). The logical implication of the court's definition of "business" was that corporate officers should be able to sue for the loss of their jobs, in contrast to prior cases that had reached the contrary conclusion by their construction of the statutory term "property." *See id.*

172. *Id.* at 419-20; *see supra* notes 169, 155-60 and accompanying text.

173. In fact, each rule originally had been formally derived with reference to the statutory language or an imputed Congressional intention to make an exception. *See supra* notes 143-45 and accompanying text.

Following *Roseland*, other opinions allowing suits by sales agents and salesmen also avoided addressing the tension between those holdings and the corporate-officer rule, and between the respective interpretive conventions.[174] More generally, the scope and authority of the concrete, "ordinary" meaning was evidently unclear: where should the language be treated as concrete, and where should that meaning yield by analogy to the rules regarding stockholders, creditors, officers, and directors?

An uneasy impasse between these two main approaches to standing analysis endured throughout the 1940s. The proponents of the ordinary meaning of the statute never rediscovered the ordinary-meaning origins of the burgeoning category rules. An awareness of those origins might have enabled them to reassert the previous concept of the ordinary meaning of the statute as a unifying general theory for antitrust standing analysis and to insist that new exclusions must meet the same tests. The proponents of the category rules, on the other hand, had not yet fabricated their general theory, the idea that these rules were all subsumed under a direct-injury principle.[175]

B. Other Standing Requirements

Within the same period, other decisions simply ignored the ordinary-meaning precedents of 1890-1914, invoking neither the results nor the reasoning of that period's case law. Two standing requirements applicable to all plaintiffs were proposed: direct injury and injury distinguishable from that of the community. Neither innovation gained substantial acceptance by 1950.

1. A Proposed Direct-Injury Standing Requirement

This section will examine the degree of support for a direct-injury standing rule during the period from 1915 to 1950. The above discussion has shown the conventional view that antitrust standing analysis always has contained a direct-injury requirement[176] to be

---

174. *See* Vines v. General Outdoor Advertising Co., 171 F.2d 487, 491 (2d Cir. 1948) (salesman stated claim for lost salary against employer who had withdrawn employment after entering into agreement violating antitrust laws); McWhirter v. Monroe Calculating Mach. Co., 76 F. Supp. 456, 460 (W.D. Mo. 1948) (district manager, paid on percentage commission, stated claim against rival company for primary-line price discrimination reducing his company's sales); Klein v. Sales Builders, Inc., 1950-1951 Trade Cas. (CCH) ¶ 62,600 (N.D. Ill. Apr. 3, 1950). That loss of employment could constitute injury to business or property supporting a claim for treble damages was also implicit in Gardella v. Chandler, 172 F.2d 402 (2d Cir. 1949).

175. *See infra* notes 257-61 and accompanying text.

176. *See supra* note 5 and accompanying text.

incorrect for the period 1890-1914.[177] The courts of 1915-1950 also generally declined to recognize the direct-injury rule as authoritative.

The first step is to look among the standing decisions of 1915-1950 for citations to prior direct-injury holdings. Of the opinions of 1890-1914, only *Loeb v. Eastman Kodak Co.*[178] contains language that could suggest a generalized direct-injury standing requirement. No court, however, cited *Loeb* for its direct-injury point for the next forty years.[179] During this period two district courts in the same circuit as *Loeb* overruled motions to dismiss antitrust claims brought by a lessor for damages resulting from harm to a lessee.[180] These courts must have either understood the direct-injury language in *Loeb* to apply only to the shareholder-creditor context,[181] or not considered it authoritative outside that context.

A few years after *Loeb*, an unambiguous direct-injury standing rule was stated in *United Copper Securities Co. v. Amalgamated Copper Co.*[182] Reversing the dismissal of a complaint, the Second Circuit declared (evidently without prompting from counsel) that under section 7

> [t]he person injured must be engaged in a business directly, or at least not remotely, affected by the conspiracy complained of. One who had rented offices to corporations absorbed by an illegal combination could not recover for losing them as tenants, nor a lawyer regularly retained for losing them as clients.[183]

Plaintiffs' assignors had been injured through harm to corporations in which they owned stock; however, because the defendants had

---

177. *See supra* notes 68-74, 106-37 and accompanying text.

178. 183 F. 704 (3d Cir. 1910). *See supra* notes 118-19 and accompanying text.

179. The next decision to cite *Loeb* for its direct-injury reference was Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).

180. *See* East Orange Amusement Co. v. Vitagraph Inc., Trade Reg. Rep. (CCH) ¶ 52,965 (D.N.J. June 24, 1943); United Exhibitors, Inc. v. Twentieth Century Fox Film Distrib. Corp., 31 F. Supp. 316 (W.D. Pa. 1940). In *East Orange Amusement Co.*, the defendant asserted that "the plaintiff, as the owner and not the operator of the theater, was not a person 'injured in his business or property' within the meaning of the statute." *East Orange Amusement Co.*, Trade Reg. Rep. (CCH) 53,566 (D.N.J. June 24, 1943). In *United Exhibitors, Inc.*, one of the plaintiffs was the lessee of a theater, and the others were owners of the same theater; the defendants' motion to dismiss did not even raise the indirectness issue. *United Exhibitors, Inc.*, 31 F. Supp. 316.

181. *See supra* notes 120-31 and accompanying text.

182. 232 F. 574 (2d Cir. 1916).

183. *Id.* at 577. The defendants' briefs on appeal made no reference to this purported requirement. Their objection to plaintiffs' standing to seek treble damages relied on the rule that injuries to a corporation can be redressed only by the corporation, not by stockholders or bondholders. *See* Brief for Amalgamated Copper Co. at 26-28, United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574 (2d Cir. 1916) (No. 120); Brief for Frederick Lewisohn at 7-8, 31, *id.*

intended to harm them individually, this was deemed sufficiently direct to support recovery.[184] The court cited no authority for its rule.[185]

Certainly in *United Copper Securities Co.*, if not beforehand, it is explicitly declared that only "directly" affected plaintiffs are entitled to sue for treble damages.[186] This purported rule enjoyed little acceptance, perhaps because it was such a transparent innovation. From the issuance of *United Copper Securities Co.* through 1950, only one decision invoked its direct-injury standing requirement.[187] In this case, *Seaboard Terminals Corp. v. Standard Oil Co.*,[188] the court dismissed the claim of a lessor damaged through harm to a lessee because the injuries were "not alleged to have been intended or designed as part of the conspiracy but only to have been a result thereof."[189] The dismissed lessor later successfully reentered the case, however, after simply amending its complaint to allege that defendants had intended to injure the lessor as well as the lessee,[190] indicating the narrowness of the *Seaboard* holding and of the constraint that court perceived in the *United Copper Securities Co.* precedent. Even this district court within the Second Circuit ignored the Second Circuit's remark in *United Copper Securities Co.* that a landlord could not seek treble damages for the loss of a tenant.[191]

---

184. *United Copper Sec. Co.*, 232 F. at 577.

> [W]e think the allegation of a conspiracy to destroy certain copper companies, for instance, the United Copper Company . . . was properly pleaded as proof of the conspiracy whereby the plaintiff's assignors were injured, notwithstanding that they were interested as stockholders of the companies and could not recover damages for corporate injuries.

Id. at 576.

185. *See id.* at 576-77.

186. *See id.* at 576, 578.

187. *See* Seaboard Terminals Corp. v. Standard Oil Co., Trade Reg. Rep. (CCH) ¶ 25,013 (S.D.N.Y. Dec. 16, 1936).

188. *Id.* Two related plaintiffs sought treble damages. *Id.* Seaboard Midland was a jobber of gasoline allegedly driven out of business by defendants. *Id.* Seaboard Midland had leased transshipping facilities from its parent corporation, Seaboard Terminals, and Seaboard Terminals claimed as damages the profits it lost from leasing those facilities to Seaboard Midland. *Id.*

189. *Id.* at 25,057 (relying on United Copper Securities Co. v. Almagamated Copper Co., 232 F. 574, 577 (2d Cir. 1916)). Only one other decision from this period asserted as a generalized standing rule that defendant must have intended to harm plaintiff. *See* Mid-West Theatres Co. v. Co-Operative Theatres, 43 F. Supp. 216, 220 (E.D. Mich. 1941).

190. *See* Seaboard Terminals Corp. v. Standard Oil Co., 24 F. Supp. 1018 (S.D.N.Y. 1938). In its earlier order of dismissal, the district court had thrown out a broad hint: "I need not decide whether . . . Seaboard [Terminals] could recover treble damages, if the charge had been that defendants' alleged conspiracy . . . had been aimed at putting Seaboard . . . together with Midland out of business." *Seaboard Terminals Corp.*, Trade Reg. Rep. (CCH) ¶ 25,013 at 25,057 (S.D.N.Y. Dec. 16, 1936).

191. *United Copper Securities Co.*, 232 F. at 577.

An examination of two decisions long cited as examples of the direct-injury rule's operation[192] further rebuts the conventional historical representation that the rule was authoritative before the 1950s.[193] A close reading reveals that both decisions strenuously avoided stating such a rule. This is especially meaningful because both arose in the Second Circuit after the *United Copper Securities Co. v. Amalgamated Copper Co.*[194] decision.

In *Gerli v. Silk Association of America*,[195] plaintiff alleged injury as an officer, controlling stockholder, and creditor of a corporation harmed by defendant's conspiracy.[196] In rejecting plaintiff's claim as a stockholder, the court simply stated that only a corporation could recover damages for injury to the corporation's business, and cited two opinions in which the rejection of plaintiff's claim had been based purely on this well-established rule of corporations law.[197]

Given the distinctive legal relationship existing between a corporation and its stockholders and creditors,[198] the denial of plaintiff's standing as a stockholder and as a creditor[199] indicated nothing about which groups of antitrust plaintiffs outside the corporate context might lack standing. The court failed to invoke a general direct-injury standing requirement, and its failure to cite *United Copper Securities Co.*[200] on this point implied rejection of that opinion's assertion of such a requirement. The plaintiff's claims for lost salary as an officer and employee were dismissed not because of indirectness but on the grounds that these were not harms to business or property.[201]

In *Westmoreland Asbestos Co. v. Johns-Manville Corp.*,[202] the court dismissed the claims of two plaintiffs whose claims derived from their relation to two corporations that had been eliminated as

---

192. *See* Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 533-34 (1983); Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).

193. *See supra* note 5 and accompanying text.

194. 232 F. 574 (2d Cir. 1916).

195. 36 F.2d 959 (S.D.N.Y. 1929).

196. *Id.*

197. *Id.* at 960 (citing Green v. Victor Talking Mach. Co., 24 F.2d 378 (2d Cir. 1928) and Ames v. American Tel. & Tel. Co., 166 F. 820 (C.C.D. Mass. 1909)).

198. *See supra* notes 113, 126, and accompanying text.

199. Gerli v. Silk Ass'n of Am., 36 F.2d 959, 960 (S.D.N.Y. 1929). The court rejected plaintiff's claim as a creditor with the terse observation that "this is not in form or substance a judgment creditor's suit." *Id.*

200. 232 F. 574 (2d Cir. 1916). *Gerli* cited *United Copper Sec. Co.* but not for its direct-injury rule. *Gerli*, 36 F.2d at 961.

201. *Gerli*, 36 F.2d at 960-61.

202. 30 F. Supp. 389 (S.D.N.Y. 1939), *aff'd per curiam*, 113 F.2d 114 (2d Cir. 1940).

competitors of the defendants.[203] Kuehn Inc. had held real estate subject to a mortgage; it had rented this real estate to the other two corporations and used the rental payments to pay carrying charges on the mortgage.[204] When the alleged conspiracy drove the renters out of business, the loss of these rent payments resulted in the foreclosure of the mortgages, and the lessor corporation sought damages for the loss resulting from the foreclosure.[205] William Kuehn, an officer and the sole stockholder of all three corporations, had used his salaries and dividends to pay carrying charges on other mortgaged real estate.[206] The loss of these salaries and dividends also resulted in foreclosure on the real estate, and Kuehn sought damages for the loss resulting from the foreclosure.[207]

The court invoked the simple category rules stated in *Gerli* in rejecting the individual's claims as a shareholder and officer of the ruined corporations.[208] In contrast, the only explanation the court gave for denying the standing of Kuehn Inc. as landlord of the more immediate targets of the conspiracy was that "the basis of the relief sought against the defendants is the foreclosure of mortgages by independent third persons" and that this was "more remote" than damages held to be unrecoverable in prior cases.[209] It was the consequences following after loss of the rent, not the lost rent itself, that the court classified as too remote to support recovery.[210]

The precise character of the remoteness concept alluded to in *Westmoreland* is unclear. Plainly, however, it is not that of *United Copper Securities Co.*, which declares categorically that landlords cannot recover treble damages for losing tenants.[211] *Westmoreland* fails to refer to this statement or resolve the issue in those straightforward categorical terms.[212] *United Copper Securities Co.* also indicates that those whom a violator intended to harm are "directly"

---

203. *Id.* at 390.
204. *Id.* at 390-91.
205. *Id.* at 391.
206. *Id.*
207. *Westmoreland Asbestos Co.*, 30 F. Supp. at 391.
208. *Id.* ("[A] stockholder cannot recover for the impairment of his stock by combination in restraint of trade . . . . [L]oss of corporate office and salary incident thereto [was not injury to] business or property within the meaning and intent of the Anti-Trust laws.").
209. *Id.* Plaintiff sought to recover for the value lost in foreclosure, owing to the interruption of receipt and salary, not for the lost rent and salary themselves. Brief for Appellant at 26, Westmoreland Asbestos Co. v. Johns-Manville Corp., 113 F.2d 114 (2d Cir. 1940) (No. 391).
210. *Westmoreland Asbestos Co.*, 30 F. Supp. at 391.
211. United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574, 577 (2d Cir. 1916).
212. *See Westmoreland Asbestos Co.*, 30 F. Supp. at 391.

injured.[213] In *Westmoreland*, plaintiffs alleged defendants drove the other two corporations out of business to injure both the individual and the lessor corporation, yet the court dismissed plaintiffs without acknowledging the issue.[214] The Second Circuit's affirmance of the district court decision[215] thus implied that the authority of *United Copper Securities Co.* was much depleted.

Another decision from the Second Circuit shows an even plainer disregard for the statement in *United Copper Securities Co.* that a landlord could not sue for damages incurred through injury to a tenant.[216] In *Camrel Co. v. Paramount Film Distributing Corp.*,[217] the court relied on the language of section 4 of the Clayton Act in overruling defendants' objection that the landlord's injury was too remote to permit recovery.[218]

Directness is applied as a standing criterion in *McAbee v. Pure Carbonic Co. of America*,[219] but in a limited way that undercuts rather than supports the *United Copper Securities Co.* direct-injury rule. Plaintiff claimed that defendants' discrimination in price against plaintiff's customers had injured him.[220] The court held that although plaintiff's business felt "a repercussion of a wrong practiced upon his customers," he could not recover because his injury was "secondary," "indirect[]," and "incidental."[221] The court inferred these standing principles, however, from section 2 of the Clayton Act, rather than the treble-damage provisions of the Clayton or Sherman Antitrust Acts, and suggested plaintiff could recover for such indirect injury if it arose from a Sherman Antitrust Act violation.[222] This dictum again indicates that the direct-injury requirement of *United*

---

213. *United Copper Sec. Co.*, 232 F. at 576-77. This is why the demurrer was overruled in that case. *See supra* notes 182-85 and accompanying text.

214. *Westmoreland Asbestos Co.*, 30 F. Supp. at 390-91.

215. Westmoreland Asbestos Co. v. Johns-Manville Corp., 113 F.2d 114 (2d Cir. 1940), *aff'g per curiam* 30 F. Supp. 389 (S.D.N.Y. 1939).

216. *United Copper Sec. Co.*, 232 F. at 577.

217. 1944-1945 Trade Cas. (CCH) ¶ 57,233 (S.D.N.Y. Mar. 30, 1944). Plaintiff owned a theatre, which it leased to an operator. *Id.* at 57,326. Plaintiff alleged that because of defendants' conspiracy, "competition by the operator of plaintiff's theatre" had been curtailed. *Id.* The lessor sought damages for the reduced "rental and market value of its theatre," and for the loss of "additional income from its lease under the percentage provisions thereof." *Id.*

218. *Id.* at 57,326.

219. 1932-1939 Trade Cas. (CCH) ¶ 55,052 (S.D.N.Y. Apr. 9, 1934).

220. *Id.* at 151.

221. *Id.* According to the court, if this price discrimination were illegal under the Clayton Act, the "immediate victims" could sue. *Id.*

222. *Id.* at 151-52. "The remedy, if any, that is open for the protection of plaintiff against the acts of defendant, is to be found in the Sherman Anti-Trust Act." *Id.* The court dismissed only the Clayton Act cause of action. *See McAbee*, 1932-1939 Trade Cas. (CCH) ¶ 55,052 at 152.

Copper Securities Co.[223] was not deemed authoritative even in its own circuit.

Further evidence of the courts' resistance to the direct-injury rule between 1915 and 1950 can be found in other decisions that reject[224] or ignore[225] standing objections based on indirectness of injury and in decisions that use ordinary-meaning reasoning in resolving standing objections.[226] Except for the handful of direct-injury decisions noted above,[227] references to directness of injury in treble-damage decisions of this period do not refer to standing.[228]

---

223. 232 F. 574, 577 (2d Cir. 1916), *aff'd*, 244 U.S. 261 (1917).

224. In Dowd v. United Mine Workers of Am., 235 F. 1, 8 (8th Cir. 1916), *cert. denied*, 242 U.S. 653 (1917), the defendant union argued that any damage to certain plaintiff coal companies was "indirect, incidental, and too remote to entitle them to recover." *Id.* at 8. The court rejected this argument, relying on the language of § 7. *Id.* at 9.

225. *See* Charles A. Ramsay Co. v. Associated Bill Posters, 260 U.S. 501 (1923). Plaintiffs, appealing dismissal of their complaint, argued that they satisfied the direct-injury standing rule of United Copper Securities Co. v. Amalgamated Copper Co., 232 F. 574, 577 (2d Cir. 1916). *See* Brief for Plaintiffs in Error at 20, 23-24, *Charles A. Ramsay Co.*, 260 U.S. 501 (Oct. Term 1922, Nos. 100-01). The Court ignored this argument, however, and addressed the § 7 point very simply: "as a result of the defendants' unlawful acts," plaintiffs were harmed; "the statute has been violated and plaintiffs' business has suffered." *Charles A. Ramsay Co.*, 260 U.S. at 511.

See also Louisiana Farmers' Protective Union, Inc. v. Great Atlantic & Pacific Tea Co. of America, 131 F.2d 419 (8th Cir. 1942), *rev'g* 40 F. Supp. 897 (E.D. Ark. 1941). The district court based its dismissal on non-standing grounds, but noted with approval defendants' objection that "plaintiffs' assignors, being neither competitors of, nor purchasers from, the defendants, are not within the class of persons . . . entitled to relief under the Clayton Act as amended by the Robinson-Patman Price-Discrimination Act." 40 F. Supp. at 916. In reversing, the Eighth Circuit ignored defendants' objections regarding standing, implicitly overruling them. 131 F.2d 419.

226. *See supra* note 163 and accompanying text.

227. *See supra* notes 188-90 and accompanying text.

228. One context in which courts often used a direct-indirect distinction was the jurisdictional requirement that a violation sufficiently affect interstate commerce. *E.g.*, 1 HARRY A. TOULMIN, A TREATISE ON THE ANTITRUST LAWS §§ 8.9, 8.24 (1949 & Supp. 1980). In some cases, the requisite 'direct' impact on interstate commerce was shown only through the impact of defendants' conduct on the plaintiff itself. *E.g.*, *Charles A. Ramsay Co.*, 260 U.S. at 511; *Dowd*, 235 F. at 8-9; *see also* Albert Pick-Barth Co. v. Mitchell Woodbury Corp., 57 F.2d 96, 99 (1st Cir.) (plaintiff need not prove that "its losses resulted directly from a suppression of its interstate trade"; it was sufficient "if they flowed from any act of the defendants in furtherance of an unlawful combination"), *cert. denied*, 286 U.S. 552 (1932).

Other references to a direct-injury requirement in treble-damage opinions of this period merely refer to plaintiff's obligation to allege and prove by nonspeculative evidence that plaintiff was injured "by reason of" defendant's violation; that defendant's violation was demonstrably the cause-in-fact of plaintiff's injury. *See* Comment, *Antitrust Enforcement by Private Parties: Analysis of Developments in the Treble Damage Suit*, 61 YALE L.J. 1010, 1017 (1952) ("Previously, courts

2. Injury Separate From That of the Public

Another generalized treble-damage standing requirement occasionally asserted was that plaintiff's injury must be separate from that common to the public.[229] In practical terms, this was not much of a restriction. It would have no consequence in the bulk of antitrust damage actions, nor was it instrumental in dismissing any actions. This assertion appears in cases in which plaintiffs had utterly failed to prove (or allege) any injury to themselves whatsoever and was probably intended only to emphasize that private plaintiffs had no standing to sue for treble damages simply as members of an injured economy.[230] No effort was made to derive this standing requirement from the language of the treble-damage provision.

---

frequently . . . stated conclusions of causal sufficiency in 'direct-indirect' terminology."). This sense of directness is often indicated by the context. *E.g.*, Momand v. Universal Film Exchs., Inc., 172 F.2d 37, 43 (1st Cir.) *cert. denied*, 336 U.S. 967 (1948); American Sea Green Slate Co. v. O'Halloran, 229 F. 77, 79 (2d Cir. 1915); Thomason v. United Gas Pub. Servs. Co., 8 F. Supp. 14, 18 (W.D. La. 1933).

Treble-damage decisions from this period also occasionally state that plaintiff must show "proximate" injury. Here, too, the context of most such statements indicates that the court was referring merely to plaintiff's obligation to establish by nonspeculative evidence that defendant's violation was the cause-in-fact of plaintiff's injury. *E.g.*, Hart v. B.F. Keith Vaudeville Exch., 12 F.2d 341, 345 (2d Cir.), *cert. denied*, 273 U.S. 703 (1926); Jack v. Armour & Co., 291 F. 741, 745 (8th Cir. 1923); Twin Ports Oil Co. v. Pure Oil Co., 46 F. Supp. 149, 153 (D. Minn. 1942); *see* Mason City Tent & Awning Co. v. Clapper, 144 F. Supp. 754, 765-66 (W.D. Mo. 1956).

In a few decisions referring to direct or proximate harm, the context does not affirmatively indicate the sense in which this terminology was intended. *E.g.*, Kellogg Co. v. National Biscuit Co., 38 F. Supp. 643 (D.N.J. 1941). There are good reasons, however, not to interpret such remarks as stating a standing requirement. Reference to directness of injury occurred widely in addressing impact on interstate commerce and the causal sufficiency of plaintiff's allegations of injury. In contrast, the directness concept was not widely used during this period to refer to issues of standing, and would not have been assumed to refer to standing, without more. One would expect, accordingly, that a court wishing to address a standing issue would denote this either by citing precedent with a clear-cut standing analysis, or by articulating its point more explicitly than by merely referring to direct or proximate harm. Even the later courts that took liberties with precedent in trying to assemble a pedigree for a direct-injury standing rule did not cite for this purpose any decisions not identified above as standing decisions. *E.g.*, Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 n.1 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952); *see infra* notes 260-61, 279, 355, and accompanying text.

229. *See* Beegle v. Thompson, 138 F.2d 875, 881 (7th Cir. 1943), *cert. denied*, 322 U.S. 743 (1944); McJunkin v. Richfield Oil Corp., 33 F. Supp. 466, 468 (N.D. Cal. 1940); Ebeling v. Foster & Kleiser Co., 12 F. Supp. 489, 490 (W.D. Wash. 1935); TOULMIN, *supra* note 228, at § 20.12; *see also supra* notes 101-02 and accompanying text; *cf.* Wheeler-Stenzel Co. v. National Window Glass Jobbers' Ass'n, 152 F. 864, 874 (3d Cir. 1907) (notes common-law principle that "every man has a right to insist that no provision of any law shall be violated, so as to work peculiar harm to him").

230. *See, e.g.*, *Beegle*, 138 F.2d at 881.

C. *Explaining the Loss of a General Theory of Antitrust Standing*

The most profound change in treble-damages standing doctrine between 1915 and 1950 was the transformation of the ordinary-meaning interpretive convention. The doctrine changed from a broadly acknowledged general principle capable of explaining all standing case law (either by the statutory language or by a discrete exception founded in congressional intent) to an approach invoked only by some courts, with even those courts evidently unsure of its scope or authority in relation to other contemporary approaches.[231] By the 1940s even the proponents of an ordinary-meaning approach to standing analysis did not view it as a unifying principle that could account for the established category rules.[232] How could this transformation have come about?

One hypothesis might be that adherence to the ordinary-meaning convention diminished after 1915 because of the challenge posed by the direct-injury rule of standing proposed in *United Copper Securities Co. v. Amalgamated Copper Co.*[233] The citation history of that decision shows, however, that the direct-injury rule was virtually a dead letter through 1950,[234] and the reasoning in the ordinary-meaning decisions of the 1940s does not indicate that the direct-injury rule was a source of concern.[235] The decline of the ordinary-meaning approach to antitrust standing doctrine was probably not inspired by dissatisfaction with the content of the holdings reached in the early ordinary-meaning decisions. The diminished stature of the ordinary-meaning convention is visible even among decisions that use it to explain why new types of plaintiffs should have standing.[236]

Instead, this fundamental change was most likely an unintended result of a simple but far-reaching change in form: the recharacterization of the original ordinary-meaning holdings as rules governing concrete categories of plaintiffs.[237] The rules barring concrete categories of plaintiffs originated in cases decided according to the ordinary-meaning approach, and the category rules initially relied for their authority on those original decisions.[238] Between 1915 and 1950, though, these category rules became the most familiar propositions

---

231. *See supra* notes 68-74, 81-87, 163-74 and accompanying text.

232. *See, e.g.*, Roseland v. Phister Mfg. Co., 125 F.2d 417, 419 (7th Cir. 1942); *see also supra* notes 169-74 and accompanying text.

233. 232 F. 574 (2d Cir. 1916); *see supra* notes 182-85 and accompanying text.

234. *See supra* notes 188-90 and accompanying text.

235. *See supra* notes 164-74 and accompanying text.

236. *See, e.g., Roseland*, 125 F.2d 417; *see also supra* notes 163-74 and accompanying text.

237. *See, e.g.*, Ames v. American Tel. & Tel. Co., 166 F. 820, 822-23 (C.C.D. Mass. 1909); *see also supra* notes 143-45 and accompanying text.

238. *See supra* notes 143-47 and accompanying text.

of antitrust standing doctrine. Courts began to view these rules as having derived their authority from judicial pronouncement rather than from the words of the statute, and consequently viewed them as incompatible with the ordinary meaning of the statute.[239] In this setting the ordinary-meaning convention no longer appeared to be a unifying general theory of antitrust standing. To support this hypothesis, two questions need to be addressed. First, how was the original ordinary-meaning justification for the concrete standing rules forgotten? Second, once forgotten, why was it not rediscovered?

The ordinary-meaning justification for the category rules seems to have been forgotten through a conjunction of factors. The idea of a general ordinary-meaning approach to the statute was not controversial originally, and the only place it was explicitly memorialized was in the original decisions.[240] In the period following the original decisions, the sense of the statute's ordinary meaning as providing a unifying general theory was never disputed, but it was never mentioned either.[241] As later decisions recited rules founded originally in ordinary-meaning reasoning (the stockholder rule, for example), the original decisions were denoted by at most a citation; later, perhaps not at all.[242]

The transformation of holdings into rules failed to attract notice. Because the rules were being applied to similar fact patterns, they were neutral in substantive impact.[243] For years after the original decisions, the outcomes reached under the prohibitory category rules could have been reconciled with the ordinary-meaning interpretive convention.[244] The existence of an earlier rule, not its etiology, was what concerned later courts. An earlier decision's reasoning was superfluous to the practical task at hand; the use of rules as a substitute would not by itself imply repudiation of the ordinary-meaning convention. Nor would rule formation by itself indicate a step toward a unitary direct-injury rule; the pertinent decisions seem to have tried to avoid precisely that.[245] By the 1940s, the original statutory-construction basis of the category rules, long since omitted from mention, seems to have been forgotten.[246] In 1914 courts would

---

239. *See supra* notes 144-47.

240. *See supra* notes 74, 80-87 and accompanying text.

241. *See, e.g.*, Corey v. Independent Ice Co., 207 F. 459, 460 (D. Mass. 1913); *see also supra* notes 144-45 and accompanying text.

242. *See, e.g.*, *Corey*, 207 F. 459; *see also supra* notes 144-45 and accompanying text.

243. *See supra* notes 143-53 and accompanying text.

244. *See supra* notes 143-53 and accompanying text.

245. *See* Westmoreland Asbestos Co. v. Johns-Manville Corp., 113 F.2d 114 (2d Cir. 1940), *aff'g per curiam* 30 F. Supp. 389 (S.D.N.Y. 1939); *see also supra* notes 192-211 and accompanying text.

246. *See supra* notes 169-74 and accompanying text.

have understood the decisions denying standing to stockholders, creditors, and corporate officers as instances of an ordinary-meaning approach to the words of the statute. By the 1940s, courts understood these holdings as free-standing rules, with no clear relation to the statute.

The second part of the question is why the original ordinary-meaning justification for the category rules, once forgotten, was not rediscovered, perhaps by a plaintiff arguing against a standing objection or a court overruling a standing objection. This may seem counterintuitive: after all, the pertinent federal reports were on the shelves of any law library. But several factors help explain why the ordinary-meaning roots of antitrust standing doctrine were not rediscovered. From the vantage point of the 1940s, nothing in antitrust standing case law would have suggested the prior existence of a general unifying theory. The outcomes were becoming increasingly hard to reconcile, and the decisions also displayed disparate forms of reasoning in their holdings.[247] For several decades courts had not mentioned or alluded to the notion of standing case law having any systematic relation to the statute; this notion had disappeared.

Large parts of antitrust standing doctrine appeared to be judge-made and were accepted as such.[248] By the 1940s the concrete rules denying standing to stockholders, creditors, officers, and directors were the most familiar propositions in antitrust standing case law, and they no longer bore traces of their origin in the ordinary-meaning approach.[249] The formal innovation of replacing exegesis with rules had been used for so long that it was taken for granted by the time its substantive impact was manifested: reasoning by analogy from the rules, without justification by the words of the statute or congressional intent.[250]

Stumbling across the ordinary-meaning principle merely by reading one of the original decisions underlying the concrete rules also would be unlikely. No single decision from the early period pointed out that all contemporary standing cases fit within the ordinary-meaning convention, which governed statutory construction generally and had no special relevance to standing issues. The unifying ordinary-meaning pattern was to some extent implicit in any single decision. Decades later the essential points of unity would not be easily discernible unless a reader already knew what to look for. Rediscovering the original ordinary-meaning justification uniting the

---

247. *See supra* notes 144-71, 176-91, 219-22 and accompanying text.

248. *See supra* notes 144-61, 182-91 and accompanying text.

249. *See supra* notes 143-45 and accompanying text. Some courts stated these rules without even citing precedent for them. *See, e.g.*, East Orange Amusement Co. v. Vitagraph Inc., Trade Reg. Rep. (CCH) ¶ 52,965 (D.N.J. June 24, 1943).

250. *See supra* notes 148-62 and accompanying text.

category rules probably would have required systematic historical investigation. The origins of antitrust standing doctrine, however, failed to arouse anyone's curiosity.

III. 1951-1975: INNOVATION

A. The Rise of a General Direct-Injury Rule

We have seen that the universally applicable direct-injury standing requirement asserted in *United Copper Securities Co.*[251] in 1916 did not achieve any substantial support by 1950.[252] In 1950 antitrust standing doctrine was conventionally viewed as a collection of rules permitting or barring certain discrete categories of plaintiffs.[253] In the 1950s, however, courts abruptly began to subscribe to a general direct-injury standing requirement. By 1960 this revolutionary innovation had become commonplace in antitrust standing doctrine.

The new campaign to promote a general direct-injury standing requirement for antitrust plaintiffs began with *Conference of Studio Unions v. Loew's Inc.*[254] Plaintiffs—an association of labor unions, the constituent unions, and individual members—alleged an antitrust conspiracy between major motion picture studios and a rival union, aimed at eliminating the plaintiffs and a group of minor motion picture studios.[255] In affirming dismissal of the complaint, the Ninth Circuit stated two related general rules governing antitrust standing.[256]

First, a plaintiff must be directly rather than incidentally injured, in the sense that the violation was "directed at" the plaintiff.[257] If plaintiff's injury were merely "incidental to the accomplishment of the illegal object," the "windfall of treble damages" would be withheld.[258] The court justified its direct-injury rule by reference to precedent:

> It has been held that shareholders, creditors, directors and officers of corporations injured by monopolistic practices of competitors cannot recover their individual losses. The reasoning of the courts in [these cases] is that the conspiracy to restrain competition was directed at the corporation and the damage suffered by the plaintiff was merely incidental.

---

251. 232 F. 574 (2d Cir. 1916).
252. *See supra* notes 188-226 and accompanying text.
253. *See, e.g.*, Coast v. Hunt Oil Co., 96 F. Supp. 53, 58-60 (W.D. La. 1951), *aff'd*, 195 F.2d 870 (5th Cir.), *cert. denied*, 344 U.S. 836 (1952).
254. 193 F.2d 51 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).
255. *Id.* at 53-54.
256. *Id.* at 54-55.
257. *Id.*
258. *Id.*

> . . . The cited cases . . . illustrate the rule that persons incidentally injured by a conspiracy cannot sue. . . . The fact that their injury was incidental was controlling.[259]

This historical representation distorts the reasoning of the cited decisions[260] and ignores much contrary authority.[261]

---

259. *Conference of Studio Unions*, 193 F.2d at 54 & n.1 (citing Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910); Gerli v. Silk Ass'n of Am., 36 F.2d 959 (S.D.N.Y. 1929); Corey v. Boston Ice Co., 207 F. 465 (D. Mass. 1913)). The footnote continues: "Nor can a landlord who lost a tenant, or a regularly retained lawyer whose services are no longer necessary, bring suit." 193 F.2d at 54 n.1 (citing Westmoreland Asbestos Co. v. Johns-Manville Corp., 30 F. Supp. 389, 391 (S.D.N.Y. 1939), *aff'd per curiam*, 113 F.2d 114 (2d Cir. 1940); United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574, 577 (2d Cir. 1916)).

260. The outcomes of the decisions cited can be explained with reference to a direct-injury principle, but the court's invocation of the "reasoning" of these decisions as supporting a direct-injury "rule," *Conference of Studio Unions*, 193 F.2d at 54 & n.1, was for the most part false.

In Corey v. Boston Ice Co., 207 F. 465 (D. Mass. 1913), the court dismissed claims for lost salary brought by officers and directors, based on the premise that this was not harm to business or property. *Id.* at 466. This decision was an instance of ordinary-meaning reasoning. *See supra* notes 135-37 and accompanying text.

Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910), referred to "indirect" injury, in holding that shareholders and creditors could not recover for losses flowing from injury to the corporation. *Id.* at 709. The main reason given for this holding, however, was that Congress must have intended to incorporate these rules in § 7 of the Sherman Antitrust Act because they were so well established in the contemporary common law of corporations. *Id.* at 708-09; *see supra* notes 114-17 and accompanying text.

Gerli v. Silk Ass'n of America, 36 F.2d 959 (S.D.N.Y. 1929), held that shareholders, creditors and officers could not recover treble damages for damage flowing from injury to the corporation. *Id.* at 960-61. The dismissal of the shareholder and creditor claims was stated very narrowly, however, based on the authority of prior cases that had derived this rule from the law of corporations. *Id.* Dismissal of the claim for lost salary as an officer had nothing to do with directness; it was based on the reasoning that this was not harm to business or property. *Id.*; *see supra* notes 195-201 and accompanying text.

For the proposition that "a landlord who lost a tenant" cannot sue, the court in *Conference of Studio Unions* cites *Westmoreland Asbestos Co.*, 30 F. Supp. 389, and *United Copper Sec. Co.*, 232 F. 574. *Conference of Studio Unions*, 193 F.2d at 54 n.1. In *Westmoreland* a landlord's claim was dismissed, but not on such categorical grounds; the holding was stated very narrowly. *Westmoreland Asbestos Co.*, 113 F.2d 114; *see supra* notes 208-11 and accompanying text. Only the dictum in *United Copper Sec.*, 232 F. at 577, supports the statement. *See supra* notes 182-84 and accompanying text.

261. The "rule that persons incidentally injured by a conspiracy cannot sue," *Conference of Studio Unions*, 193 F.2d at 54, is contradicted by earlier case law in several respects. Such a rule was clearly incompatible with the ordinary-meaning approach that dominated early antitrust standing analysis, *see supra* notes 68-72, 80-87, and accompanying text, and was specifically rejected during that period. *See supra* notes 91-92 and accompanying text. The direct-injury rule enjoyed little support between 1916 and 1950. *See supra* notes 188-223 and accompanying text. It is also tacitly contradicted by decisions allowing suit by plaintiffs incidentally injured by a

The court's second standing requirement was that a plaintiff must be "within that area of the economy which is endangered by a breakdown of competitive conditions in a particular industry."[262]

> Such a construction is in accordance with the basic and underlying purposes of the anti-trust laws to preserve competition and to protect the consumer. Recovery and damages under the anti-trust law is [*sic*] available to those who have been directly injured by the lessening of competition . . . .[263]

The court failed to explain why the stated purposes of the antitrust laws require this limitation on the scope of the remedy.[264] The invocation of the overall purposes of the antitrust laws in general rather than the purposes of the treble-damage section marked a new step in distancing formal standing analysis from the text of the treble-damage provision in the statute. Given the direct-injury rule articulated in the same opinion, the "lessening of competition" test seems to have been intended to exclude the injuries of labor plaintiffs occurring in the context of an antitrust violation, if such injuries can be traced to labor union rivalry.[265]

Within six years of the Ninth Circuit's *Conference of Studio Unions* decision, some version of a generally applicable direct-injury requirement had been adopted by the Courts of Appeals for the

---

conspiracy intended to injure defendants' competitors. *See, e.g.*, Thomsen v. Union Castle Mail S.S. Co., 166 F. 251 (2d Cir. 1908), *aff'd sub nom.* Thomsen v. Cayser, 243 U.S. 66 (1917); H.B. Marienelli, Ltd. v. United Booking Offices of America, 227 F. 165 (S.D.N.Y. 1914). Cases from the decade before *Conference of Studio Unions* in which incidentally injured plaintiffs had been permitted to sue include Roseland v. Phister Mfg. Co., 125 F.2d 417 (7th Cir. 1942); Klein v. Sales Builders, Inc., 1950-1951 Trade Cas. (CCH) ¶ 62,600 (N.D. Ill. Apr. 3, 1950); McWhirter v. Monroe Calculating Mach. Co., 76 F. Supp. 456 (W.D. Mo. 1948); Camrel Co. v. Paramount Film Distrib. Corp., 1944-1945 Trade Cas. (CCH) ¶ 57,233 (S.D.N.Y. Mar. 30, 1944); and East Orange Amusement Co. v. Vitagraph Inc., Trade Reg. Rep. (CCH) ¶ 52,965 (D.N.J. June 24, 1943). *Conference of Studio Unions* is the original source of the conventional history of antitrust standing doctrine, which the present Article criticizes.

262. *Conference of Studio Unions*, 193 F.2d at 54-55.

263. *Id.* at 55.

264. *Id.* at 54-55.

265. By "competition" the court seems to mean "commercial competition." *See id.* at 54. The rival unions were not in commercial competition; hence, corresponding injuries were not remediable under the antitrust laws. *See id.* ("insofar as the conspiracy was to restrain trade by destroying competitors it was directed at the [minor studios], and the alleged damage the [plaintiffs] suffered therefrom was incidental to the accomplishment of the illegal object"). In the context of the case, the two standing rules fulfilled complementary roles. To the extent an antitrust violation was directed at the minor studios, the labor plaintiffs' harm was not direct. To the extent a violation was directed at the labor plaintiffs, their injury was not related to commercial competition.

First,[266] Second,[267] and Third[268] Circuits. Other courts hesitated to adopt the sweeping generality of the direct-injury rule,[269] but none opposed it.

Courts formulated the direct-injury requirement variously. Inconsistency arose not only between circuits,[270] but among opinions of the courts of appeals in individual circuits,[271] and even among district courts and their respective circuit courts of appeal.[272] The middle ground in this loose consensus was that a plaintiff lacked standing unless its injury was "direct," meaning there was no intervening victim, or the violation was "directed" at the plaintiff, meaning the plaintiff was an intended victim.[273] Despite the purported connection

---

266. *See* Miley v. John Hancock Mut. Life Ins. Co., 148 F. Supp. 299 (D. Mass.), *aff'd*, 242 F.2d 758 (1st Cir. 1957), *cert. denied*, 355 U.S. 828 (1957); Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907 (D. Mass. 1956).

267. *See* Productive Inventions, Inc. v. Trico Prod. Corp., 224 F.2d 678 (2d Cir. 1955), *cert. denied*, 350 U.S. 936 (1956).

268. *See* Melrose Realty Co. v. Loew's, Inc., 234 F.2d 518 (3d Cir.), *cert. denied*, 352 U.S. 890 (1956); Harrison v. Paramount Pictures, Inc., 115 F. Supp. 312 (E.D. Pa. 1953), *aff'd*, 211 F.2d 405 (3d Cir. 1954), *cert. denied*, 348 U.S. 828 (1954).

269. *See* Congress Bldg. Corp. v. Loew's, Inc., 246 F.2d 587, 590-91 (7th Cir. 1957); Martens v. Barrett, 245 F.2d 844, 846 (5th Cir. 1957); Peter v. Western Newspaper Union, 200 F.2d 867 (5th Cir. 1953).

270. *Compare, e.g.*, Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951) (persons "incidentally" injured by a conspiracy directed at another lack standing), *cert. denied*, 342 U.S. 919 (1952) *with* Productive Inventions, Inc. v. Trico Prod. Corp., 224 F.2d 678, 679 (2d Cir. 1955) ("only those at whom the violation is directly aimed, *or who have been directly harmed* may recover") (emphasis added), *cert. denied*, 350 U.S. 936 (1956).

271. *Compare, e.g., Conference of Studio Unions*, 193 F.2d at 54 ("persons incidentally injured by a conspiracy cannot sue") *with* Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 362 (9th Cir. 1955) (standing depends on whether plaintiff was "within the 'target area'" of defendants' illegal practices) *and* Steiner v. 20th Century-Fox Film Corp., 232 F.2d 190, 193 (9th Cir. 1956) (refers to directness of injury, but cites only cases from other circuits; ignores *Conference of Studio Unions* and *Karseal*).

272. *Compare* Harrison v. Paramount Pictures, Inc., 115 F. Supp. 312, 317 (E.D. Pa. 1953) ("It is not possible to formulate any general rule by which to determine what injuries are too remote . . . ."), *aff'd*, 211 F.2d 405 (3d Cir.), *cert. denied*, 348 U.S. 828 (1954) *with* Gomberg v. Midvale Co., 157 F. Supp. 132, 142 (E.D. Pa. 1955) (statute requires "injury to the economy of the plaintiff, by virtue of restrictions of trade or something that proximately follows from it [sic], in the competitive field in which it is engaged") (citing only precedents from outside Third Circuit) *and* Rossi v. McCloskey & Co., 149 F. Supp. 638, 640 (E.D. Pa. 1957) ("Injury which is merely a collateral effect of illegal restraint upon competition is not compensable . . . .").

273. *See Productive Inventions, Inc.*, 224 F.2d at 679. The most expansive version of the direct-injury rule articulated in the 1950s was that of *Karseal Corp.*, 221 F.2d 358. The narrowest version would probably be that in *Harrison*, 115 F. Supp. 312; *see* Melrose Realty Co. v. Loew's Inc., 234 F.2d 518, 519-20 (3d Cir.) (Biggs, C.J., dissenting), *cert. denied*, 352 U.S. 890 (1956).

with the early case law, many of the new direct-injury decisions denied standing to plaintiffs who seemed to have met the direct-injury requirements of *United Copper Securities Co.*[274]

The direct-injury standing rule of the 1950s continued to gain wider acceptance through the early 1960s.[275] By this time the directness concept was increasingly conveyed by the terms "target area" and "proximate cause,"[276] but the results reached were within the range of variation found in the direct-injury opinions of the 1950s.

The courts that followed the Ninth Circuit's direct-injury rule seem to have been aware it was an innovation. The assertion in *Conference of Studio Unions* that the direct-injury rule had been a well-established requirement since early in the century[277] is recognizably inaccurate;[278] it was not repeated by contemporary courts.[279] A number of courts overtly grounded the direct-injury standing requirement in propositions of policy,[280] and some courts attributed this

---

274. 232 F. 574, 576-77 (2d Cir. 1916); *see supra* notes 182-84 and accompanying text. A number of owner-lessors were denied standing to sue for damages resulting from conspiracies that included the owners' lessees. *See Melrose Realty Co.*, 234 F.2d at 519-20 (Biggs, C.J., dissenting); Lieberthal v. North Country Lanes, Inc., 221 F. Supp. 685, 690 (S.D.N.Y. 1963), *aff'd on other grounds*, 332 F.2d 269 (2d Cir. 1964); Skouras Theatres Corp. v. Radio-Keith-Orpheum Corp., 193 F. Supp. 401, 407 (S.D.N.Y. 1961); *Harrison*, 115 F. Supp. at 317. No intervening victims existed between such plaintiffs and the conspirators, the lessors' injuries were clearly to be expected, and the lessors suffered the most substantially from the restraints; these plaintiffs would therefore seem to be "directly, or at least not remotely, affected by the conspiracy complained of." *United Copper Sec.*, 232 F. at 577. *See* Steiner v. 20th Century-Fox Film Corp., 232 F.2d 190, 193 (9th Cir. 1956) (owner of theatre was directly injured by conspiracy among lessee and other parties).

275. *See, e.g.*, Lieberthal v. North County Lawns, Inc., 221 F. Supp. 685, 689 (S.D.N.Y. 1963), *aff'd on other grounds*, 332 F.2d 269 (2d Cir. 1964); *Skouras Theatres Corp.*, 193 F. Supp. at 407.

276. *See* Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty and Sur. Co., 382 F.2d 925, 928 (10th Cir. 1967).

277. Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).

278. *See* Comment, *Antitrust Enforcement By Private Parties: Analysis of Developments in the Treble Damage Suit*, *supra* note 228 at 1019-20 ("lack of precedent" for the direct-injury rule noted); Taggart Whipple, *Two Aspects of Plaintiffs' Treble Damage Suits: Class Actions; Person Injured and Standing to Sue*, 8 A.B.A. ANTITRUST SECTION 27, 38 & n.62 (1956) (notes "contrary" earlier decisions); Marvin Grove, Comment, *Means of Determining When Treble Damages Are Recoverable For a Loss Due To a Violation of the Antitrust Laws*, 46 CAL. L. REV. 447, 448-54 (1958) (cites decisions that allow "indirect" losses); Berger & Bernstein, *supra* note 5, at 818 ("In the early 1950s . . . the direct injury rule was a rule of uncertain authority.").

279. Those that came closest were Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 363 (9th Cir. 1955), and Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956).

280. *See* Congress Bldg. Corp. v. Loew's, Inc., 246 F.2d 587, 591 (7th Cir.

policy-based reasoning to others that did not articulate it.[281] Some of the decisions that ascribe their direct-injury requirement to precedent rather than to deliberate choice nonetheless indicate the latter by their reformulation of the terms of the requirement.[282]

It was a matter of consensus in the 1950s that standing doctrine was a subject for judicial lawmaking rather than statutory exegesis. Courts that applied the direct-injury rule invoked policy, precedent, or both; none purported to derive their holdings from an ordinary-meaning gloss on the statutory language of the treble-damage sections or from legislative history pertaining to these sections.[283] Some courts overtly contrasted the breadth of the statute's language and the narrower direct-injury rule, without attempting to link the doctrine to the statute.[284] The direct-injury decisions treated the broad language of the treble-damage sections as at most the source of a policy, to be weighed along with other policies.[285] The term "standing" first

---

1957) (suggests that practical distinctions between types of antitrust plaintiffs merit differing standing rules); Melrose Realty Co. v. Loew's, Inc., 234 F.2d 518, 519 (3d Cir.) (per curiam) ("we think it is a sound rule"), *cert. denied*, 352 U.S. 890 (1956); *Snow Crest Beverages, Inc.*, 147 F. Supp. at 909 ("policy considerations" support direct-injury rule); Harrison v. Paramount Pictures, Inc., 115 F. Supp. 312, 317 (E.D. Pa. 1953) (industry would be burdened unless indirect injuries were excluded; statute should be construed to avoid "unreasonable results"), *aff'd per curiam*, 211 F.2d 405 (3d Cir.), *cert. denied*, 348 U.S. 828 (1954); *cf.* Fanchon & Marco, Inc. v. Paramount Pictures, Inc., 202 F.2d 731, 735 (2d Cir. 1953) (court allowed stockholder's derivative suit for treble damages: "we do not believe sound policy consistent with greater restriction on the right to treble damages"). The validity and sufficiency of such policy propositions were assumed, not addressed. *See Fanchon*, 202 F.2d at 735.

281. *E.g.*, *Congress Bldg. Corp.*, 246 F.2d at 591 (attributes the courts' standing restrictions to "the ground, more or less expressed, that in view of the severity of the penalty a line must be drawn"); *Snow Crest Beverages, Inc.*, 147 F. Supp. 907 (attributes the courts' "narrow construction of Section 4 of the Clayton Act" to "policy considerations"); *see also* Whipple, *supra* note 278, at 34 ("there has been a marked divergence in judicial views as to the desirable scope of the statutory language").

282. *See supra* notes 257-60, 266-72 and accompanying text.

283. *See supra* notes 270-72, 280-82 and accompanying text.

284. *See, e.g.*, *Congress Bldg. Corp.*, 246 F.2d at 591 ("Notwithstanding the broad language of Section 4, which literally, at least, appears to give the plaintiff herein a cause of action, the courts have narrowed the apparent scope of this statute . . . ."); Schwartz v. Broadcast Music, Inc., 180 F. Supp. 322, 327 (S.D.N.Y. 1959); *Snow Crest Beverages, Inc.*, 147 F. Supp. at 909; *Harrison*, 115 F. Supp. at 317 (treble-damages section "should not be literally construed if unreasonable results would be reached by doing so").

285. *See, e.g.*, *Congress Bldg. Corp.*, 246 F.2d at 591; Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 365 (9th Cir. 1955). In *Congress Bldg. Corp.* the court made an interesting comment regarding an earlier standing decision using an ordinary-meaning approach to the statute, Camrel Co. v. Paramount Film Distrib. Corp., 1944-1945 Trade Cas. (CCH) ¶ 57,233 (S.D.N.Y. Mar. 30, 1944). *See Congress Bldg. Corp.*, 246 F.2d at 588-89. The court commented that *Camrel* had

was used generally to refer to questions of eligibility to sue for treble damages during this period:[286] the new usage seems to express the modern understanding that these questions are a matter for doctrinal elaboration rather than actual interpretation of the statute. Commentators accepted the "extrastatutory" nature of the direct-injury rule;[287] critical evaluation centered on whether it appropriately accommodated relevant policy objectives.[288]

B. A Public-Injury Standing Requirement

A second major instance of revolutionary development in treble-damages standing doctrine in the 1950s was a requirement that the defendants' violation of the antitrust laws must have injured the public as well as the plaintiff.[289] This requirement resembled definitions of the substantive violation under the Sherman Antitrust Act; it would be redundant if courts accepted "public injury" as an essential element of antitrust violations.[290] Because of its resemblance to definitions of substantive violation, this standing requirement is often hard to identify with certainty in particular cases[291] except by

---

not given "substantial treatment to the problem," apparently referring to its lack of policy-based reasoning. *Id.*

When Congressional policy was invoked in standing decisions, it tended to be either policy attributed to the antitrust laws as a whole, or policy relating to some overall purpose of providing a private remedy. *See, e.g., Karseal Corp.*, 221 F.2d at 365.

286. *See, e.g.*, Productive Inventions, Inc. v. Trico Prod. Corp., 224 F.2d 678, 679 (2d Cir. 1955), *cert. denied*, 350 U.S. 936 (1956); Whipple, *supra* note 278.

287. *See* David Westfall, *Landowners' Third-Party Damage Suits Under the Sherman Antitrust Act*, 72 HARV. L. REV. 305, 306 (1958); Whipple, *supra* note 278, at 34 ("there has been a marked divergence in judicial views as to the desirable scope of the statutory language"); Grove, Comment, *supra* note 278, at 448 ("the decisions have created a distinction between 'direct' and 'indirect' losses").

288. *E.g.*, Westfall, *supra* note 287, at 306; Comment, *Should "Injury" in Treble Damage Suits Be Redefined?*, 51 NW. U. L. REV. 141, 146-47 (1956); Recent Case, 104 U. PA. L. REV. 543, 545 (1956) ("Any justifiable determination of who may sue under section 4 must be based on fulfilling the purposes of that section."); Comment, *Third-Party Recovery For Injury To Economic Interests—A Common-Law Problem In Interpreting the Antitrust Laws*, *supra* note 170, at 712 ("Apart from the cases, it would seem that assistance in determining the scope of permissible plaintiffs might be derived from the policy underlying treble damages.").

289. *See infra* notes 292-93 and accompanying text.

290. *See* 2 PHILLIP AREEDA & DONALD F. TURNER, ANTITRUST LAW § 331c, at 151 (1978) ("The public injury [standing] issue may reflect confusion concerning the definition of the substantive offense.").

291. An example is provided by the lower court decisions in Klor's v. Broadway-Hale Stores, Inc., 255 F.2d 214 (9th Cir. 1958), *rev'd*, 359 U.S. 207 (1959). The district court's order dismissed the count in question because plaintiff's damages were not "caused by a public wrong proscribed by the [Sherman] Act." 255 F.2d

implication: reference to public injury as an element of a private action rather than as an element of any Sherman Antitrust Act violation, whether in a private action or a government action.[292] Some courts, however, more unequivocally stated the public-injury test as a standing requirement.[293]

The case for the public-injury standing requirement relied mainly on Supreme Court dicta of dubious relevance to standing doctrine.[294]

---

at 220 n.17, 221. The plaintiff's specification of error in the court of appeals characterized the district court's dismissal as turning on a standing ground: "[T]he complaint had failed to allege and plaintiff had failed to prove, as an additional element, a 'public injury.'" *Id.* at 221. Defendants disagreed with this characterization of the district court's order, *id.*, and the court of appeals affirmed the dismissal on the ground that no antitrust violation was proved. *Id.* at 235.

292. *E.g.*, Radiant Burners, Inc. v. Peoples Gas, Light & Coke Co., 273 F.2d 196, 200 (7th Cir. 1959) (in absence of per se violation, private action exists "only under circumstances where there is such general injury to the competitive process that the public at large suffers economic harm"), *rev'd per curiam*, 364 U.S. 656 (1961); Donlan v. Carvel, 209 F. Supp. 829, 831 (D. Md. 1962) ("except for *per se* violations of the Sherman Antitrust Act, a plaintiff to recover must allege and prove that the public has been adversely affected"); Schwing Motor Co. v. Hudson Sales Corp., 138 F. Supp. 899, 907 (D. Md.) (complaint alleged no "injury to the public sufficient to support an action for treble damages"), *aff'd per curiam on other grounds*, 239 F.2d 176 (4th Cir. 1956), *cert. denied*, 355 U.S. 823 (1957); *cf.* Richfield Oil Corp. v. Karseal Corp., 271 F.2d 709, 727 (9th Cir. 1959) (plaintiff must prove "*public injury*, i.e. a violation of the anti-trust laws"), *cert. denied*, 361 U.S. 961 (1960).

In some of the first decisions later cited as leading cases for the public-injury standing doctrine, the references to public injury actually refer to the substantive scope of the Sherman Antitrust Act. *See, e.g.*, Shotkin v. General Elec. Co., 171 F.2d 236, 238 (10th Cir. 1948). On the question of private recovery, *Shotkin* merely states that plaintiff must show that the defendants "did in fact violate the Act with resulting damages to plaintiff." *Id.* at 239.

293. *See, e.g.*, *In re* McConnell, 370 U.S. 230, 231 (1962) ("erroneous holding" of trial court requiring proof of public injury noted); Nelligan v. Ford Motor Co., 262 F.2d 556, 559 (4th Cir. 1959) (complaint must allege facts that show an antitrust violation "and from which an inference of public injury may reasonably be extracted"); Parmelee Transp. Co. v. Keeshin, 186 F. Supp. 533, 541 (N.D. Ill. 1960) (proof of public injury essential in private treble damage suit), *aff'd on other grounds*, 292 F.2d 794 (7th Cir.), *cert. denied*, 368 U.S. 944 (1961); Admiral Theatre Corp. v. Paramount Film Distrib. Corp., 140 F. Supp. 686, 695 (D. Neb. 1955) ("it must be shown that not only the individual plaintiff has sustained damages because of an alleged violation, but that public rights have also been violated"); Slick Airways, Inc. v. American Airlines, Inc., 15 F.R.D. 175, 182 (D.N.J. 1954) (plaintiff must show that "defendants violated the antitrust laws to the damage of the plaintiff and the public as well").

294. Proponents of the public-injury standing requirement commonly invoked remarks from two Supreme Court decisions, neither of which addressed standing requirements. *See* D.R. Wilder Mfg. v. Corn Prods. Ref. Co., 236 U.S. 165 (1915); Apex Hosiery Co. v. Leader, 310 U.S. 469 (1940). In *D.R. Wilder Mfg.*, the Court stated that the "prohibitions" of the Sherman Antitrust Act were intended to "prevent not the mere injury to an individual which would arise from the doing of the prohibited acts, but the harm to the general public," and that "not only the

It ignored the contrary implications of many early cases[295] and the Supreme Court's pronouncements defining a broad scope for the treble-damage remedy.[296] Many courts supported the idea of a public-injury rule; others opposed it.[297] The Supreme Court seemingly repudiated the public-injury rule in 1961,[298] but some courts continued to apply the rule in cases not presenting per se illegality.[299]

---

prohibitions of the statute but the remedies which it provided were co-extensive with such conceptions." *D.R. Wilder Mfg.*, 235 U.S. at 174. The issue in this case was whether the defendant in a contract action could assert as a defense to an otherwise legal contract that the plaintiff corporation "had no legal existence" because it had been organized in violation of the Sherman Antitrust Act. *Id.* at 175-76. In *Apex Hosiery Co.*, the Court stated that the Sherman Antitrust Act was meant to prevent "restraints to free competition in business and commercial transactions which tended to restrict production, raise prices or otherwise control the market to the detriment of purchasers or consumers . . . , all of which had come to be regarded as a special form of public injury." *Apex Hosiery Co.*, 310 U.S. at 493. The issue under discussion was the extent to which the Sherman Antitrust Act prohibited conspiracies by employees to restrain the interstate trade of their employers. *Id.* at 502-03.

295. *See, e.g.*, Montague & Co. v. Lowry, 193 U.S. 38, 46 (1904) (recovery by private plaintiff with slight share of trade upheld); *see also* Eastern States Retail Lumber Dealers' Ass'n v. United States, 234 U.S. 600, 614 (1914) (concerted refusal to deal violates antitrust laws "if the result be hurtful to *the public or to the individual* against whom the concerted action is directed") (emphasis added). It was never claimed that the public-injury standing requirement was instrumental in any of the early case law.

296. *See, e.g.*, Radovich v. National Football League, 352 U.S. 445, 454 (1957) ("In the face of such a policy [favoring private antitrust plaintiffs] this Court should not add requirements to burden the private litigant beyond what is specifically set forth by Congress . . . ."); Mandeville Island Farms, Inc. v. American Crystal Sugar Co., 334 U.S. 219, 236 (1948) ("The [Sherman Antitrust] Act is comprehensive in its terms and coverage, protecting all who are made victims of the forbidden practices by whomever they may be perpetrated.").

297. *See* Rogers v. Douglas Tobacco Bd. of Trade, Inc., 266 F.2d 636, 644 (5th Cir.) ("specific public injury" need not be proved "before a private person can recover"), *cert. denied*, 361 U.S. 833 (1959); Bankers Life & Casualty Co. v. Larson, 257 F.2d 377, 381 (5th Cir.) (complaint alleged "at once both public and private injury"), *cert. denied*, 358 U.S. 879 (1958); Professional & Business Men's Life Ins. Co. v. Bankers Life Co., 163 F. Supp. 274, 285 (D. Mont. 1958) ("the complaint herein charges a per se violation of the Sherman Antitrust Act, and therefore, one which . . . Congress has, by legislative fiat, determined to be injurious to the public"); William Goldman Theatres, Inc. v. Twentieth-Century Fox Film Corp., 151 F. Supp. 840, 843 (E.D. Pa. 1957) ("when competition is restrained by means declared unlawful by the antitrust laws, the interest of the public is harmed"; no further allegation of public injury required).

298. *See* Radiant Burners, Inc. v. Peoples Gas, Light & Coke Co., 364 U.S. 656, 659-60 (1961) (per curiam); *see also In re* McConnell, 370 U.S. 230, 231 (1962) ("we have held that the right of recovery of a plaintiff in a treble damage antitrust case does not depend at all on proving an economic injury to the public") (citing *Radiant Burners, Inc.*, 364 U.S. at 660, and Klor's, Inc. v. Broadway-Hale Stores, Inc., 359 U.S. 207 (1959)).

299. *See, e.g.*, Donlan v. Carvel, 209 F. Supp. 829, 831 (D. Md. 1962). *Contra*, Harrison v. Prather, 435 F.2d 1168, 1176 (5th Cir. 1970), *cert. denied*, 404

C. *The Impetus for Innovation*

Interesting questions are posed by the sudden rise to prominence of the direct-injury and public-injury standing doctrines in the 1950s. What inspired such a burst of judicial innovation? Why did the direct-injury doctrine achieve much wider support than the public-injury doctrine?

The direct-injury and public-injury doctrines were parallel in many ways. Both were highly restrictive standing doctrines abruptly embraced by lower federal courts in the same decade, sixty years after the enactment of the treble-damages remedy. Both denied standing to types of plaintiffs previously allowed to sue.[300] Neither was compelled by principles of statutory construction or authority of precedent, and the early precedents that were cited on behalf of these doctrines had been around for years without having inspired such a response.[301] The parallel and abrupt emergence of the two restrictive doctrines after so many years of private antitrust litigation therefore seems likely to reflect some common stimulus.

Decisions expressly invoking policy grounds for the direct-injury standing requirement typically assert that because the trebling of damages is a drastic remedy, defendants should be protected from excessive cumulative damages.[302] In determining other points bearing on the treble-damage remedy, similarly, the courts also asserted the objective of protecting defendants from excessive damages.[303] A sus-

---

U.S. 829 (1971); Highland Supply Corp. v. Reynolds Metals Co., 238 F. Supp. 561, 563 (E.D. Mo. 1965); Switzer Bros., Inc. v. Locklin, 297 F.2d 39, 47 (7th Cir. 1961), *cert. denied*, 369 U.S. 851 (1962); Syracuse Broadcasting Co. v. Newhouse, 295 F.2d 269, 276-77 (2d Cir. 1961).

300. *See supra* notes 261, 294, and accompanying text.

301. For example, direct-injury opinions commonly cited Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910); public-injury opinions commonly cited D.R. Wilder Mfg. Co. v. Corn Prods. Ref. Co., 236 U.S. 165 (1915). *See supra* notes 257-60, 294, and accompanying text.

302. *See* Klor's, Inc. v. Broadway-Hale Stores, Inc., 255 F.2d 214, 217 (9th Cir. 1958), *rev'd*, 359 U.S. 207 (1959); Congress Bldg. Corp. v. Loew's, Inc., 246 F.2d 587, 591 (7th Cir. 1957) ("the courts have narrowed the apparent scope of this statute on the ground, more or less expressed, that in view of the severity of the penalty a line must be drawn"); Harrison v. Paramount Pictures, Inc., 115 F. Supp. 312, 317 (E.D. Pa. 1953) (industry would be burdened unless indirect injuries excluded; statute should be construed to avoid "unreasonable results"), *aff'd per curiam*, 211 F.2d 405 (3d Cir.), *cert. denied*, 348 U.S. 828 (1954). In Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907 (D. Mass. 1956), the court asserted that if indirectly injured plaintiffs were allowed to sue, "businessmen would be subjected to liabilities of indefinable scope for conduct already subject to drastic private remedies." *Id.* at 909. The court conceded that this was "a position narrower than that often applied in nonstatutory tort cases and in cases where plaintiffs are not allowed a multiple recovery." *Id.*

303. *See, e.g.*, Paramount Film Distrib. Corp. v. Applebaum, 217 F.2d 101,

picion arises, however, that this expressed concern for defendants was in part a rationalization for another objective.

The stated concern for defendants implies apprehension that a particular violator would be sued by too many plaintiffs. This goal of avoiding excessive treble-damage recovery would be at stake, however, only where multiple plaintiffs had sued or were likely to sue a defendant based on the same violation, and where some marginal plaintiff's ability to recover would turn on standing doctrine. Yet in some of the very decisions that pioneered the new standing restrictions, these conditions are not present. Some of these decisions state nonstanding grounds for dismissal, so the standing point is unnecessary to the outcome;[304] some eliminate the only plaintiffs who had filed or seemed likely to file actions based on the violation in question.[305]

What caused the courts abruptly to devise drastic new standing restrictions for private antitrust plaintiffs? Probably the critical impetus was the courts' recently manifested alarm over the rising burdens associated with antitrust litigation. Private antitrust filings were of a new magnitude, and antitrust litigation was increasingly singled out for contributing to the increase in federal court caseloads both in numbers and in relative scale and complexity.[306] Federal judges publicly expressed impatience with antitrust litigation; their attitude was characterized by some as "hostility."[307] Judges com-

---

105 (5th Cir. 1954) (a private treble-damage action "partakes of the nature of a criminal charge," so that "the proof should be stronger than in an ordinary civil action"), *cert. denied*, 349 U.S. 961 (1955); Allgair v. Glenmore Distilleries Co., 91 F. Supp. 93, 97 (S.D.N.Y. 1950) ("The extraordinary remedy of triple damages . . . requires the closest scrutiny of the transaction . . . ."); Westor Theatres v. Warner Bros. Pictures, 41 F. Supp. 757, 762 (D.N.J. 1941) (because treble-damage remedy is "unusual" and "drastic," statutory provisions must be strictly construed).

304. *See* Nelligan v. Ford Motor Co., 262 F.2d 556 (4th Cir. 1959); Radiant Burners, Inc. v. Peoples Gas, Light & Coke Co., 273 F.2d 196 (7th Cir. 1959), *rev'd*, 364 U.S. 656 (1961); Conference of Studio Unions v. Loew's Inc., 193 F.2d 51 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952); Miley v. John Hancock Mut. Life Ins. Co., 148 F. Supp. 299 (D. Mass), *aff'd*, 242 F.2d 758 (1st Cir.), *cert. denied*, 355 U.S. 828 (1957); Schwing Motor Co. v. Hudson Sales Corp., 138 F. Supp. 899 (D. Md.), *aff'd per curiam on other grounds*, 239 F.2d 176 (4th Cir. 1956); Admiral Theatre Corp. v. Paramount Film Distrib. Corp., 140 F. Supp. 686 (D. Neb. 1955); Harrison v. Paramount Pictures, Inc., 115 F. Supp. 312 (E.D. Pa. 1953), *aff'd*, 211 F.2d 405 (3d Cir.), *cert. denied*, 348 U.S. 828 (1954).

305. *See* Melrose Realty Co. v. Loew's, Inc., 234 F.2d 518 (3d Cir.), *cert. denied*, 352 U.S. 890 (1956); *Conference of Studio Unions*, 193 F.2d 51; *Harrison v. Paramount Pictures*, 115 F. Supp. 312.

306. *See, e.g.*, John T. Chadwell & Richard W. McLaren, *The Current Status of the Antitrust Laws*, 1950 U. ILL. L. F. 491, 513-14 (1950) ("the increasing volume of all litigation in the federal courts, in general, and the increased number and complexity of antitrust cases, in particular, have clogged the dockets of the district courts").

307. *See* Eagle Lion Studios, Inc. v. Loew's, Inc., 248 F.2d 438, 451 (2d Cir.

plained that antitrust litigation, being particularly time-consuming and burdensome, posed a severe problem for the courts.[308] The Judicial Conference assembled a committee to study the problems of antitrust procedure, which found in antitrust litigation an "acute major problem in the current administration of justice" that might "threaten the judicial process itself."[309] Private antitrust suits outnumbered government antitrust suits and were now, like government suits, considered likely to generate the dreaded "big case."[310] The

---

1957) (Clark, C.J., dissenting) (notes "developing trend in some of our trial courts of hostility toward the 'big' antitrust case"), *aff'd*, 358 U.S. 100 (1958); Lee Loevinger, *Private Action—The Strongest Pillar of Antitrust*, 3 ANTITRUST BULL. 167, 170 (1958) ("many courts have an attitude of definite hostility to private antitrust suits"). In overruling an award of attorney's fees, one court declared that "the possibility that [treble-damage litigation] might develop into a racketeering practice should not be enhanced." Milwaukee Towne Corp. v. Loew's, Inc., 190 F.2d 561, 570 (7th Cir. 1951), *cert. denied*, 342 U.S. 909 (1952).

308. Chief Judge John Clark Knox, *Response*, 1952 N.Y. ST. B.A. SEC. ANTITRUST LAW SYMP. 14, 15. Judge Knox stated:

> The thing that presently worries me is that antitrust litigations in the court over which I preside are monopolizing the time, energy and effort of judges who ought to be trying cases that have to do with the lame, the halt and the blind, and who are daily being deprived of the rights of simple justice.

Id. See also Judge Leon R. Yankwich, *"Short Cuts" in Long Cases*, 13 F.R.D. 41, 42 (1952) (antitrust litigation "notoriously" requires extensive proof, and generates "complex problems"); 1950 ANN. REP. DIRECTOR ADMIN. OFFICE U.S. COURTS 43-44 ("The burden of antitrust litigation is not so much measured by the number of cases as by the time consumed to dispose of each one."), *quoted in* Chadwell & McLaren, *supra* note 306, at 514 n.106. *See also* Breck P. McAllister, *The Big Case: Procedural Problems in Antitrust Litigation*, 64 HARV. L. REV. 27, 28 (1950) (antitrust litigation presents "a serious problem of judicial administration"; if not dealt with, "the processes of antitrust litigation may break down of their own weight"); Whipple, *supra* note 278, at 41 (antitrust litigation "stigmatized by its elephantine bulk and boundless prolixity"); Wm. Dwight Whitney, *The Trial of an Anti-Trust Case*, 5 REC. ASS'N B. CITY N.Y. 449, 469 (1950) ("The present procedure in large commercial litigation has failed.").

309. JUDICIAL CONFERENCE OF THE U.S., COMMITTEE REPORT, PROCEDURE IN ANTI-TRUST AND OTHER PROTRACTED CASES, 13 F.R.D. 62, 64 (1952).

> Cases of this sort, especially in the anti-trust and patent fields, . . . are of sufficient frequency to create an acute major problem in the current administration of justice. Unnecessary consumption of time and energy, delay in disposition of disputes, and enormous expenditures of money are among the vices resulting from the circumstances described, but the principal vice is that such conditions create confusion, magnify uncertainty, multiply the possibilities of error, and otherwise tend to make less certain and more accurate the judicial determination of disputed issues. The latter vices, if permitted to continue to exist, might threaten the judicial process itself in respect to complex controversies.

Id.

310. *See* H. Templeton Brown, *Venue and Statute of Limitations*, 1952 N.Y. ST. B.A. SEC. ANTITRUST LAW SYMP. 43 ("the 'big case' is not necessarily" a government action); Whitney, *supra* note 308, at 466 ("[E]ven these Government

purported concern regarding cumulative recoveries against antitrust defendants, which did not directly complicate the judicial function, never aroused federal judges to a comparable extent.

The adoption of the direct-injury and public-injury standing requirements corresponded closely in time to these public expressions of impatience among the lower federal courts, and responded directly to the courts' perceived need for relief from this crisis. Other rules associated with judicial efforts to protect antitrust defendants from excessive recoveries, such as rules regarding the fact and quantum of damages,[311] were fact-intensive and could not be implemented until after considerable litigation, often not until after trial; they did not substantially lessen the court's workload. The distinctive contribution of standing doctrine was to eliminate entire lawsuits in the earliest stages of litigation.[312] One contemporary observer detected in current standing doctrine "judicial self-defense against multiplicity of suits."[313] The lower courts' rulings on other matters were commonly attributed to antipathy toward antitrust litigation.[314] It would be curious if

---

cases can be dwarfed by triple damages cases."). A contemporary study of antitrust litigation indicated that "the median time from filing to disposition was 23 months for government civil cases and 22 months for private civil cases, as compared with a median time of disposal of less than ten months for all tried civil cases." 1950 DIRECTOR ADMIN. OFFICE U.S. COURTS ANN. REP. 43-44, *quoted in* Chadwell & McLaren, *supra* note 306, at 514 n.106.

311. *See, e.g.*, Homer Clark, *The Treble Damage Bonanza: New Doctrines of Damages in Private Antitrust Suits*, 52 MICH. L. REV. 363, 363 (1954) ("the mandatory trebling of any recovery has generated a natural reluctance in the courts to impose prodigious damages upon violators of the act"); Loevinger, *supra* note 307, at 170 ("the decisions [in private antitrust suits] are stricter in their interpretation of the laws and more rigorous in their requirements of proof . . . than in government actions under the antitrust laws"); Westfall, *supra* note 287, at 305 (1958) ("There has been a recurring tendency to view the remedy of treble damages as a 'windfall,' to be allowed only in the rarest cases"); Comment, *Antitrust Enforcement By Private Parties: Analysis of Developments in the Treble Damage Suit*, *supra* note 228, at 1018-19 (detects continued use of antitrust damages rules to "circumvent" pertinent Supreme Court authority favorable to private plaintiffs).

312. *See* Sherman, *supra* note 5, at 376-77 & nn.16-18.

313. Whipple, *supra* note 278, at 38. The opinion in Conference of Studio Unions v. Loew's Inc., 193 F.2d 51 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952), which led the innovations of the 1950s, may have been inspired by impatience with particular plaintiffs; this was the third action based on the same underlying controversy. The other two actions had already come up on appeal; in each, dismissal had been affirmed. *See* Schatte v. International Alliance of Theatrical Stage Employees, 182 F.2d 158, 163 (9th Cir.), *cert. denied*, 340 U.S. 827 (1950); Schatte v. International Alliance of Theatrical Stage Employees, 165 F.2d 216 (9th Cir.), *cert. denied*, 334 U.S. 812 (1948).

314. *See, e.g.*, Eagle Lion Studios, Inc. v. Loew's, Inc., 248 F.2d 438 (2d Cir. 1957) (Clark, C.J., dissenting), *aff'd*, 358 U.S. 100 (1958). Chief Judge Clark observed:

> a developing trend in some of our trial courts of hostility toward the "big" antitrust case and of discovering obstacles—going even back to matters of

standing decisions failed to reflect that antipathy, when standing doctrine was so unformed and offered such a uniquely responsive solution to the problem.[315]

A second characteristic of both the direct-injury and the public-injury rules that lower federal courts of the 1950s would have found appealing was that both were relatively simple, coherent, single-factor principles of general applicability. Judges are generally said to be drawn toward "fixed and regular determinateness" of rights.[316] For judges who would have to decide standing objections, both of the new rules offered a welcome improvement over the multiplicitous, atheoretical standing case law of the 1940s.

These two objectives—reducing the number of antitrust suits and introducing doctrinal simplicity—would explain judicial support for both the direct-injury and public-injury rules of treble-damage standing. Two other criteria, however, help explain why the direct-injury doctrine achieved so much greater acceptance among the federal courts of the 1950s and early 1960s than the public-injury rule did: the latter exhibited much less continuity with precedent and a much less plausible relation to the pertinent statutory provision.

The two standing rules differed greatly in terms of apparent continuity with earlier standing case law. The pedigree for the direct-injury rule stated in *Conference of Studio Unions*[317] seemed to have been recognized at the time as overstated,[318] but it sounded plausible: at least the outcomes of the decisions cited were explainable with reference to directness of injury.[319] The rule's plausibility would have

---

> pleading and pre-trial—in the way of a free showing of the need of remedial relief. Humanly speaking we can well sympathize, for these trials are a burden, if not a bore, to a busy, overworked court.

Id. at 451. *See also* Lee Loevinger, *Handling a Plaintiff 's Antitrust Damage Suit*, 4 ANTITRUST BULL. 29, 30 (1959) ("there is a noticeable tendency on the part of at least some courts to apply the law somewhat more narrowly and to be somewhat stricter in demanding proof in a private [antitrust] action than in a government action"). *See supra* note 310 and accompanying text.

315. *See* Berger & Bernstein, *supra* note 5, at 812 ("Underlying these standing decisions may be a concern of a completely different kind: denial of standing may be perceived as a means of easing the administrative burdens of antitrust actions"). To take an example that was prominent at the time, the Oil Jobber cases of the 1940s ultimately resulted in no recovery against defendants among over forty lawsuits. *See* Comment, *Antitrust Enforcement by Private Parties: Analysis of Developments in the Private Treble Damage Suit*, *supra* note 228, at 1046-47. Many cases were lost on the issue of damages; generally this would not be decided at the outset of the case. *See id.*

316. 2 MAX WEBER, ECONOMY AND SOCIETY 875 (Guenther Roth & Claus Wittich eds. 1968).

317. *Conference of Studio Unions*, 193 F.2d at 54.

318. *See supra* notes 278-79 and accompanying text.

319. *See* Gerli v. Silk Ass'n of Am., 36 F.2d 959 (S.D.N.Y. 1929); Corey v. Boston Ice Co., 207 F. 465 (D. Mass. 1913); Loeb v. Eastman Kodak Co., 183 F. 704 (3d Cir. 1910).

increased with time because the direct-injury opinions of the 1950s generated a canon of favorable precedents. While these opinions tended to ignore precedents that failed to fit the direct-injury rule,[320] the rule could actually explain the outcomes of most of the well-known decisions, and no one put forward any other way of harmonizing the case law.[321] Given the disparity of case law in the 1930s and 1940s, by the 1950s courts may well have thought that the soundness of the various holdings should be tested by whether they fit a unitary principle that explained the bulk of the results. The direct-injury rule was the only such unifying rationale available.

The proposed public-injury standing rule, in contrast, had no plausible doctrinal background. The only support for this rule among early case law came from decisions addressing aspects of antitrust law other than standing;[322] no early cases turn on the public-injury rule.[323]

The two rules also differed in their relation to the underlying treble-damage provision. Both rules contradicted the ordinary meaning of the statute ("[a]ny person injured by reason of any . . ."),[324] but they differed with respect to the plausibility of the respective justifications for doing so. Because direct-injury language could be found among early standing decisions,[325] Congress arguably had acquiesced by now in a direct-injury standing rule.[326] No such argument could be made on behalf of the public-injury standing rule. The two rules also differed in apparent impact on the commonly-cited deterrent purpose of the treble-damage remedy.[327] The direct-injury concept implied that when an "indirectly" injured plaintiff was dismissed for lack of standing, some other more directly injured plaintiff could bring suit for the same alleged violation; the standing dismissal would not altogether shield the alleged violator from the

---

320. *See, e.g., Conference of Studio Unions*, 193 F.2d at 54; Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 363 (9th Cir. 1955); Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956). Each of these gave a doctrinal overview that ignored such recent ordinary-meaning decisions as Roseland v. Phister Mfg., 125 F.2d 417 (7th Cir. 1942), and McWhirter v. Monroe Calculating Machine Co., 76 F. Supp. 456 (W.D. Mo. 1948).

321. *See supra* notes 172-75, 269, and accompanying text.

322. *See, e.g.*, Apex Hosiery Co. v. Leader, 310 U.S. 469, 493 (1940); D.R. Wilder Mfg. v. Corn Prods. Ref. Co., 236 U.S. 165, 174 (1915); *see also supra* note 294 and accompanying text.

323. *See supra* note 295 and accompanying text.

324. 15 U.S.C. § 15 (1973 & Supp. 1990).

325. *See, e.g.*, United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574, 577 (2d Cir. 1916).

326. *See* Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956).

327. *See, e.g.*, Kinnear-Weed Corp. v. Humble Oil & Ref. Co., 214 F.2d 891, 893 (5th Cir. 1954), *cert. denied*, 348 U.S. 912 (1955); Maltz v. Sax, 134 F.2d 2, 4 (7th Cir.), *cert. denied*, 319 U.S. 772 (1943).

treble-damage remedy. Many courts that dismissed plaintiffs on direct-injury grounds in the 1950s alluded to the existence of other entities more directly injured by the same violation than the plaintiff.[328] The public-injury rule was considerably weaker on this point. The dismissal of a plaintiff based on a public-injury standing rule often left no apparent alternative plaintiff who could sue the alleged violator, which negated the deterrent role of the private remedy.

All four of the cited objectives—reducing the number of antitrust plaintiffs, introducing doctrinal simplicity, exhibiting continuity with earlier case law, and plausibly respecting the underlying treble-damages provision—would lead courts of the 1950s to favor the direct-injury rule. The public-injury standing rule might appeal as a simple means of relieving caseloads. Its comparative weakness, however, in terms of doctrinal continuity and relation to the statute explains why the public-injury rule failed to generate a consensus among even the lower courts that felt overburdened with antitrust litigation.

D. The 1960s: A Deteriorating Consensus

In the middle and late 1960s, some courts continued to apply the direct-injury rule of standing in ways falling within the bounds of 1950s case law.[329] Other courts began to depart from that consensus, however. Some applied or redefined the direct-injury rule in ways that expanded or further restricted the scope of standing. Some challenged the direct-injury rule altogether.

1. Differentiation of Direct-Injury Concepts

Beginning in the mid-1960s courts began to use existing formulations of the direct-injury rule in ways that expanded the effective scope of standing to sue for treble damages. In some instances the expansion was minor;[330] in others it was considerable, such as allowing

---

328. *See, e.g.*, Productive Inventions, Inc. v. Trico Prods. Corp., 224 F.2d 678, 680 n.1 (2d Cir. 1955), *cert. denied*, 350 U.S. 936 (1956); Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952); *Snow Crest Beverages, Inc.*, 147 F. Supp. at 908.

329. *See, e.g.*, Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty & Sur. Cos., 382 F.2d 925, 929 (10th Cir. 1967).

330. *See, e.g.*, South Carolina Council of Milk Producers, Inc. v. Newton, 360 F.2d 414 (4th Cir.), *cert. denied*, 385 U.S. 934 (1966). Prior cases had indicated that the scope of a "target area" extended to other steps in the chain of distribution beyond the one most immediately affected. This was the case only when the same product was concerned; it did not extend to suppliers of ingredients. *See, e.g.*, Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 364 (9th Cir. 1955). In *South Carolina Council of Milk Producers* the court held that raw milk, which plaintiffs sold, was "in essentially the same form" as the processed milk sold by its customers, so plaintiffs might sue for their injury stemming from injury to their customers "within the rationale of *Karseal*." *South Carolina Council of Milk Producers*, 360 F.2d at 419.

officers of injured corporations to sue for diminished compensation.[331] In *Perkins v. Standard Oil Co.*[332] the Supreme Court disapproved a ruling denying standing to an individual owed brokerage fees, rent payments, and "other indebtedness" by corporations (owned by him) that were harmed by defendants' price discrimination at another level of distribution.[333]

Other courts in the 1960s justified expansive rulings on standing issues by substantially reformulating prior concepts of directness. This occurred most strikingly in the Ninth Circuit. In *Karseal Corp. v. Richfield Oil Corp.*,[334] the court had already put forward the "target area" test as a gloss on the direct-injury rule of *Conference of Studio Unions*.[335] The target area included all who dealt in the targeted product, not just the most immediate victim;[336] it excluded suppliers outside the product line.[337] In 1964, the court gave a greatly

---

331. *See* Data Digests, Inc. v. Standard & Poor's Corp., 43 F.R.D. 386, 387-88 (S.D.N.Y. 1967); *cf.* Schroeter v. Ralph Wilson Plastics, Inc., 49 F.R.D. 323 (S.D.N.Y. 1969) (president of defendant corporation allowed to sue for injuries suffered in his individual capacity as stockholder of defendant corporation). This contravened not only the 1950s direct-injury doctrine, but also the rule in prior standing case law. *See supra* notes 145, 169 and accompanying text.

332. 395 U.S. 642 (1969), *rev'g* 396 F.2d 809 (9th Cir. 1968).

333. *Id.* at 649-50. The Court disagreed with the Ninth Circuit's conclusion that these losses represented incidental injury:

> It is clear in this case . . . that Perkins was no mere innocent bystander; he was the principal victim of the price discrimination practiced by Standard. Since he was directly injured and was clearly entitled to bring this suit, he was entitled to present evidence of all of his losses to the jury.

Id.

This seems to have been dictum: later discussion in *Perkins* indicates the jury did not consider those claims. *See id.* at 650. Many later courts, however, read it as part of the holding. *See, e.g., In re* Multidistrict Vehicle Air Pollution M.D.L. No. 31, 481 F.2d 122, 128-29 (9th Cir.), *cert. denied*, 414 U.S. 1045 (1973).

Even the Ninth Circuit decision, while denying standing as to those injuries, had allowed plaintiff to sue for harm to "the going concern value of his interest, as owner and lessor and as prime lessee and sublessee of service stations and bulk plants," incurred through harm to corporations to which Perkins had leased these various properties. Standard Oil Co. v. Perkins, 396 F.2d 809, 815 (9th Cir. 1968), *rev'd on other grounds*, 395 U.S. 642 (1969).

334. 221 F.2d 358 (9th Cir. 1955).

335. *See id.* at 364-65; Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952). In *Karseal* a manufacturer sued for losses resulting from injury to its distributor-customers by defendant's exclusive dealing arrangement. *Karseal*, 221 F.2d at 360-61. The court might have granted standing under the *Conference of Studio Unions* rule by finding that the violation was directed at the manufacturer rather than at the distributors. However, based on inferring a constructive intent on the part of defendants to harm the competing product at all levels, the court determined that both the distributors and the manufacturer were within the "target area" of the violation. *Id.* at 364-65.

336. *See supra* notes 254-59 and accompanying text.

337. *Karseal*, 221 F.2d at 364.

expanded definition of the target area: the plaintiff had only to show "that, whether or not then known to the conspirators, plaintiff's affected operation was actually in the area which could *reasonably be foreseen to be affected by the conspiracy.*"[338] During this period the Ninth Circuit reversed standing-based dismissals of a borrower injured through a restraint injuring its lender,[339] lessors injured through restraints on their lessees,[340] and a motion picture producer whose revenues were reduced through a violation aimed elsewhere, at television stations.[341] A district court used the foreseeability test to permit standing to the officer of a corporation harmed by a violation, and the court indicated that other employees of the corporation also would be entitled to sue.[342]

Many decisions from other circuits also adopted liberalized formulations of the direct-injury or target-area tests.[343] Whether the various decisions containing more tolerant standing analyses were motivated by lessened concern for defendants, increased sympathy for plaintiffs, diminished apprehension of the burden that antitrust litigation imposed on courts, or simply increased deference to the statutory language, is not apparent from the opinions.

Not all courts were prepared to liberalize the direct-injury rule simply because the Ninth Circuit had reconsidered its earlier innovation. Some courts used the same technique of reformulating existing concepts to innovate in the opposite direction, constricting standing

---

338. Twentieth Century Fox Film Corp. v. Goldwyn, 328 F.2d 190, 220 (9th Cir.) (emphasis added), *cert. denied*, 379 U.S. 880 (1964); *see also* Hoopes v. Union Oil Co., 374 F.2d 480, 486 (9th Cir. 1967) (under the foreseeable target-area test "it was no bar to recovery . . . that appellants' injuries did not result from the allegedly illegal restraint . . . but rather from the means which [defendant] used to accomplish that restraint"); *cf.* Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956) (many plaintiffs denied standing under direct-injury test "even though as a matter of logic their losses were foreseeable").

339. *See* Harman v. Valley Nat'l Bank, 339 F.2d 564 (9th Cir. 1964).

340. *See Hoopes*, 374 F.2d at 484-85; Standard Oil Co. v. Perkins, 396 F.2d 809, 815 (9th Cir. 1968), *aff'd in relevant part, rev'd on other grounds*, 395 U.S. 642 (1969).

341. Mulvey v. Samuel Goldwyn Prods., 433 F.2d 1073, 1076 (9th Cir. 1970), *cert. denied*, 402 U.S. 923 (1971).

342. Isidor Weinstein Inv. Co. v. Hearst Corp., 303 F. Supp. 646, 650 (N.D. Cal. 1969) (complaint sufficiently alleged that "defendant could have foreseeably realized that such violation could drive certain firms out of business resulting in loss of employment to a class of people of whom plaintiff is one"); *see also* Washington v. American Pipe & Constr. Co., 280 F. Supp. 802, 807 (D. Or. 1968) (reasonable-foreseeability test supported standing of purchasers from nonconspiring sellers, where price increases allegedly resulted from defendants' price-fixing).

343. *See, e.g.*, International Ass'n of Heat & Frost Insulators & Asbestos Workers v. United Contractors Ass'n, 483 F.2d 384, 397 (3d Cir. 1973); Minnesota v. United States Steel Corp., 299 F. Supp. 596 (D. Minn. 1969); Dailey v. Quality Sch. Plan, Inc., 380 F.2d 484 (5th Cir. 1967); Epstein v. Dennison Mfg., 1966 Trade Cas. (CCH) ¶ 71,953 (S.D.N.Y. Dec. 8, 1966).

more than before.[344] A line of Second Circuit decisions furnishes the best example of this. In adopting the direct-injury rule in *Productive Inventions, Inc. v. Trico Products Corp.*,[345] the court had stated, "only those at whom the violation is directly aimed, or who have been directly harmed may recover."[346] The apparent meaning of this disjunctive phrasing was that the scope of standing included plaintiffs who were the intended victims and plaintiffs who had no intervening victim between them and the defendant's conduct. The court also cautioned that no "hard and fast rule" could be stated because "the line between direct and incidental damage is not always definable with clarity."[347]

In the late 1960s, however, the Second Circuit aggressively began to narrow the *Productive Inventions* standard. Perhaps mischievously, it chose to use the "target area" concept for this purpose.[348] Reworking both the *Productive Inventions* concept of directness and the *Karseal* concept of a target area, the court now said that the target area consisted only of the targets themselves.[349] As applied, moreover, the target area tended to consist of defendants' competitors only.[350] Because standing was still denied where another victim intervened between the defendants and the plaintiff, even if the plaintiff and defendants were competitors,[351] the alternative requirements of *Productive Inventions*[352] were effectively transformed into a double requirement: the violation must have directly injured the plaintiff and also must have been aimed at the plaintiff.[353] Even though in *Productive Inventions* the court had eschewed a bright-line rule, the

---

344. *See, e.g.*, Productive Inventions, Inc. v. Trico Prods. Corp., 224 F.2d 678, 679 (2d Cir. 1955), *cert. denied*, 350 U.S. 936 (1956) ("Those harmed only incidentally by anti-trust violations have no standing to sue . . . .").

345. *Id.*

346. *Id.* at 679.

347. *Id.* at 680.

348. *See* Calderone Enters. Corp. v. United Artists Theatre Circuit, Inc., 454 F.2d 1292, 1295 (2d Cir. 1971), *cert. denied*, 406 U.S. 930 (1972); SCM Corp. v. Radio Corp., 407 F.2d 166, 169 (2d Cir.), *cert. denied*, 395 U.S. 943 (1969).

349. *Calderone Enters. Corp.*, 454 F.2d at 1296 ("the only relevant factor is whether the plaintiff is a 'target' of the illegal activity"); *see supra* notes 335-37, 346-47 and accompanying text.

350. *See, e.g.*, GAF Corp. v. Circle Floor Co., 463 F.2d 752, 759 (2d Cir. 1972) ("only a person whose competitive business position was harmed by the *anticompetitive* effects of the alleged restraint of trade can maintain a treble damage action"), *petition for cert. dismissed*, 413 U.S. 901 (1973); *Calderone Enters. Corp.*, 454 F.2d at 1295 (target area consisted of those "against whom the conspiracy was aimed, such as a competitor of the persons sued").

351. *See, e.g.*, Billy Baxter, Inc. v. Coca-Cola Co., 431 F.2d 183, 188 (2d Cir. 1970), *cert. denied*, 401 U.S. 923 (1971).

352. 224 F.2d 678; *see supra* notes 273, 346-47 and accompanying text.

353. *See Calderone Enters. Corp.*, 454 F.2d at 1295-97.

court now observed that its "easily identifiable cutoff" possessed "the virtue of definiteness."[354]

Both the Second and Ninth Circuits, like other courts picking and choosing from various elements of direct-injury case law, purported to be acting within the bounds of precedent.[355] The direct-injury doctrine lent itself to such free reworking: the terminology was inherently imprecise and became ever more so as courts applied it in new ways. Its content always had been somewhat arbitrary. None of the direct-injury decisions of the 1950s had explained why the line between direct and indirect injury, or the boundary of a target area, should be drawn precisely one way rather than another; some of these decisions denied the standard could be precisely defined.[356] Now, pronounced variation continued to occur not only between circuits,[357] but also within individual circuits.[358] The Supreme Court steadfastly declined to resolve the confusion.[359]

Amid the proliferation of standing definitions based loosely on direct-injury precedents, more and more courts instead invoked policy as the real basis for standing rules.[360] Observers criticized rules as inevitably clumsy vehicles for implementing policy and urged courts

---

354. *Id.* at 1296 & n.2. The court's new standing test was also more restrictive than that set forth in United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574, 576-77 (2d Cir. 1916). *See supra* notes 182-84, 274, and accompanying text.

355. *See, e.g., In re* Multidistrict Vehicle Air Pollution M.D.L. No. 31, 481 F.2d 122, 127 (9th Cir.), *cert. denied*, 414 U.S. 1045 (1973); GAF Corp. v. Circle Floor Co., 463 F.2d 752, 758-59 (2d Cir. 1972), *petition for cert. dismissed*, 413 U.S. 901 (1973); *Calderone Enters. Corp.*, 454 F.2d at 1295-96.

356. *See, e.g.*, Productive Inventions, Inc. v. Trico Prods. Corp., 224 F.2d 678, 680 (2d Cir. 1955), *cert. denied*, 350 U.S. 936 (1956).

357. *See, e.g., Multidistrict Vehicle Air Pollution*, 481 F.2d at 126-28; Berger & Bernstein, *supra* note 5, at 819-24, 830-35; David L. Swiden, Note, *Standing to Sue in Private Antitrust Litigation: Circuits in Conflict*, 10 IND. L. REV. 533 (1977).

358. *See, e.g.*, Berger & Bernstein, *supra* note 5, at 833-34 & n.12; Swiden, *supra* note 357, at 541-43, 550-52; *see also* ALAN WATSON, FAILURES OF THE LEGAL IMAGINATION 139-40 (1988) ("The courts may reinterpret [a statute] away from its original purpose, but then the resulting law is as difficult to find and know as any other judge-made law.").

359. *See, e.g.*, Swiden, *supra* note 357.

360. *See, e.g., In re* Western Liquid Asphalt Cases, 487 F.2d 191, 198-200 (9th Cir. 1973), *cert. denied*, 415 U.S. 919 (1974); Billy Baxter, Inc. v. Coca-Cola Co., 431 F.2d 183, 187 (2d Cir. 1970), *cert. denied*, 401 U.S. 923 (1971); Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty & Sur. Cos., 382 F.2d 925, 929 (10th Cir. 1967); Minnesota v. United States Steel Corp., 299 F. Supp. 596, 601-02 (D. Minn. 1969); Epstein v. Dennison Mfg., 1966 Trade Cas. (CCH) ¶ 71,931, at 83,378 (S.D.N.Y. December 8, 1966).

Although some courts made conclusory remarks about congressional intent underlying the treble-damage sections of the antitrust statutes, for the most part the policies invoked in favor of restrictive standing requirements were policies of the courts. *See, e.g.*, Calderone Enters. Corp. v. United Artists Theatre Circuit, Inc., 454 F.2d 1292, 1295-96 (2d Cir. 1971), *cert. denied*, 406 U.S. 930 (1972).

to implement the relevant policies overtly instead of using rules to justify outcomes inspired by policy considerations.[361]

2. Doctrinal History and the Final Rejection of the Ordinary-Meaning Approach to the Statute

Beginning in the mid-1960s a few courts challenged the direct-injury rule and proposed that the treble-damage language of the statute simply should be applied according to its ordinary meaning for purposes of antitrust standing analysis.[362] Among the case law of the 1960s and 1970s this was by far the most fundamental challenge to the concepts of the 1950s. Not only did the ordinary-meaning approach imply a consistently broader scope of standing than the other approaches, it also rejected the legitimacy of judicial policy-making as the primary determinant of standing rules. It was the only approach to standing analysis in this period that gave primary importance to the statutory language.

The fabricated history that had accompanied the rebirth of the direct-injury rule now helped to preserve it against this attack. In

---

361. *See* Milton Handler, *The Shift from Substantive to Procedural Innovations in Antitrust Suits—the Twenty-Third Annual Antitrust Review*, 71 COLUM. L. REV. 23, 30 (1971); Berger & Bernstein, *supra* note 5, at 844-45; David Klingsberg, *Bull's Eyes and Carom Shots: Complications and Conflicts on Standing to Sue and Causation Under Section 4 of the Clayton Act*, 16 ANTITRUST BULL. 351, 369 (1971); *Calderone Enters. Corp.*, 454 F.2d at 1303 (Levet, J., dissenting).

362. Usually, this proposition was accompanied by a statement of more conventional grounds for upholding a plaintiff's standing. *See, e.g.*, Mulvey v. Samuel Goldwyn Prods., 433 F.2d 1073, 1076 (9th Cir. 1970), *cert. denied*, 402 U.S. 923 (1971); SCM Corp. v. Radio Corp. of America, 407 F.2d 166, 172 (2d Cir.) (Timbers, J., dissenting), *cert. denied*, 395 U.S. 943 (1969); Hoopes v. Union Oil Co., 374 F.2d 480, 485 (9th Cir. 1967); Harman v. Valley Nat'l Bank, 339 F.2d 564, 567 (9th Cir. 1964); Wilson v. Ringsby Truck Lines, Inc., 320 F. Supp. 699, 702-03 (D. Colo. 1970); Isidor Weinstein Inv. Co. v. Hearst Corp., 303 F. Supp. 646, 650 (N.D. Cal. 1969); Schulman v. Burlington Indus., Inc., 255 F. Supp. 847, 851 (S.D.N.Y. 1966).

An exception to the pattern of indicating alternative reasoning for upholding a plaintiff's standing is *In re* Multidistrict Motor Vehicle Air Pollution M.D.L. No. 31, 52 F.R.D. 398 (C.D. Cal. 1970), *rev'd in part*, 481 F.2d 122 (9th Cir.), *cert. denied*, 414 U.S. 1045 (1973).

> We are now concerned with the phrase "injured in his business or property *by reason of anything* forbidden in the anti-trust laws" in the light of the allegations of these complaints, rather than the traditional, legalistic approach defined by the cases cited by defendants in their motion to dismiss. Each of the plaintiffs allege injury to their respective business or property *by reason of* anti-trust violations of the defendants.

Id. at 401; *see also* Joseph L. Alioto & Peter J. Donnici, *Standing Requirements for Antitrust Plaintiffs: Judicially Created Exceptions to a Clear Statutory Policy*, 4 U.S.F. L. REV. 205 (1970); *cf.* Missouri v. Stupp Bros. Bridge & Iron Co., 248 F. Supp. 169, 173-74 (W.D. Mo. 1965).

Conference of Studio Unions, the Ninth Circuit indicated that its direct-injury rule of antitrust standing had been in force since 1910.[363] Some contemporaries, at least, appeared to have recognized the inaccuracy of this representation.[364] By the 1960s, however, the same historical account seems to have been universally accepted.[365]

It thus took only a decade for what was originally an unconvincing fabrication to become an accepted historical truth. Perhaps this astonishing shift in perspective was a natural response in the circumstances, however. The new origin-myth had been concisely and prominently stated in *Conference of Studio Unions*;[366] judges of the 1950s, who lacked any reason to discredit a rule they favored, had not contradicted it. The *Conference of Studio Unions* history was plausible on its face and was supported by some evidence: the selected canon of older cases cited in support of the direct-injury rule fit in terms of outcomes if not usually in terms of reasoning. For later readers, the direct-injury opinions of the 1950s would read as if they were continuing a conventional consensus, not inaugurating one. Earlier cases outside the selective canon were less and less likely to be noticed, especially given the relative bulk of the case law of the 1950s and 1960s. The dominance of the direct-injury concept in contemporary case law perhaps made it seem natural that this dominance had always existed.[367]

Commentators tended to focus on precisely prescribing the content of the direct-injury rule, not on establishing its legitimacy or exploring doctrinal history. Their tendency to arrange the cases by formula or outcome, rather than chronologically, reinforced a static, ahistorical picture. The conjectural (and erroneous) historical connection drawn at this time between early antitrust standing doctrine and tort doctrines of privity supported the impression that early standing doctrine was highly restrictive.[368] That discrepancies between

---

363. *See supra* note 259 and accompanying text.

364. *See supra* notes 260-61, 277-79 and accompanying text.

365. *See, e.g., Calderone Enters. Corp.*, 454 F.2d at 1298-99 (Levet, J., dissenting); Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty & Sur. Cos., 382 F.2d 925, 929 (10th Cir. 1967); Data Digests, Inc. v. Standard & Poor's Corp., 43 F.R.D. 386, 387-88 (S.D.N.Y. 1967); Klingsberg, *supra* note 361, at 352; Note, *Private Treble Damage Antitrust Suits: Measure of Damages for Destruction of All or Part of a Business*, 80 HARV. L. REV. 1566, 1571 (1967); David L. Swiden, Note, *Standing to Sue For Treble Damages Under Section 4 of the Clayton Act*, 64 COLUM. L. REV. 570, 582 (1964).

366. Conference of Studio Unions v. Loew's Inc., 193 F.2d 51 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952). The myth was repeated in Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358 (9th Cir. 1955).

367. *See supra* notes 266-69, 329-55 and accompanying text.

368. *See, e.g.*, Malamud v. Sinclair Oil Corp., 521 F.2d 1142, 1149 (6th Cir. 1975) (early antitrust standing analysis adopted privity concept); International Ass'n of Heat & Frost Insulators & Asbestos Workers v. United Contractors Ass'n, 483

the older case law and the direct-injury rule would be discounted, if they were noticed at all, is not surprising.

Those who attacked the direct-injury doctrine in the 1960s failed to realize that the historical record offered substantial grounds for attacking the false history of the direct-injury rule and for supporting the ordinary-meaning approach. These critics had every motive to explore means of undercutting the perceived legitimacy of the direct-injury rule, and the evidence for such an attack was on hand in every law library; courts still occasionally cited some of the relevant cases for other purposes. Along with everyone else, apparently, these critics of the direct-injury doctrine saw the texts of early case law through the perspective of modern doctrine, so strongly was the pattern of the direct-injury rule projected backwards onto the past.

The unquestioned representation of antitrust standing doctrine as historically coterminous with the direct-injury rule severely undermined the argument in favor of the ordinary-meaning standard in the 1960s. The failure to challenge this historical premise meant erroneously conceding a point in favor of direct-injury doctrine—its supposed longevity—that would strongly support its claim to legitimacy. The conventional historical account obscured the reality that much pre-1950s case law contradicted the direct-injury standing rule and that the case law of 1890-1914 exhibited an ordinary-meaning approach to standing issues.[369]

An informed reading of the original stockholder cases, to take an important example, would have revealed that the earliest established rule of antitrust standing case law, the rule barring stockholders from seeking damages for harm to the corporation lowering the value of their stock, could be accommodated within the ordinary-meaning approach without conceding the rest of the direct-injury doctrine.[370] Those who attacked the direct-injury doctrine as a whole without making this distinction impliedly opposed even the venerable rule against stockholder damage suits, with its protection (understandably cherished by judges) against multiple suits by stockholders whose damages could be redressed fully by suit in the corporation's name.[371]

The argument against the direct-injury standing doctrine, blinkered by the contemporary erroneous representation of doctrinal his-

---

F.2d 384, 396 (3d Cir. 1973) ("attempting to delimit the scope of the duty of care under a theory of proximate cause, courts have evolved a doctrine which holds that the defendant is liable for harm to persons who are within the 'target area'"); *Calderone Enters. Corp.*, 454 F.2d at 1302 (Levet, J., dissenting) (directness doctrine originated "in the early part of the [twentieth] century when the concept of privity dominated almost any kind of recovery"); Billy Baxter, Inc. v. Coca-Cola Co., 431 F.2d 183, 190 (2d Cir. 1970) (Waterman, J., dissenting) ("those were the days when 'privity' was king"), *cert. denied*, 401 U.S. 923 (1971).

369. *See supra* notes 68-74, 80-87, 192-228 and accompanying text.

370. *See supra* notes 80-87, 109-34 and accompanying text.

371. *See supra* note 84 and accompanying text.

tory, instead relied on two points. First, critics attacked the direct-injury doctrine as artificial, consisting merely of "judicial glosses."[372] Supporters of the direct-injury rule, however, had long conceded that it was artificial,[373] so this point was not especially telling. This artificiality was sometimes even raised in support of the direct-injury rule's legitimacy: that a doctrine departing from the statutory language had been around so long (it was asserted) without being overruled by statute meant Congress must approve of it.[374] Now, moreover, some courts further improved the purported history of the direct-injury doctrine, perhaps in response to the tardy challenge from the ordinary-meaning approach.[375]

The second argument made by supporters of the ordinary-meaning approach to standing issues was that the Supreme Court had never approved the direct-injury rule[376] and had made broad statements suggesting that no extrastatutory restrictions were appropriate.[377] The force of this premise was greatly reduced, however, by the Court's later dictum in *Hawaii v. Standard Oil Company*.[378] In holding that a state was not authorized to bring a *parens patriae* suit for injury to the state's general economy, the Court remarked in passing that

---

372. *See* Hoopes v. Union Oil Co., 374 F.2d 480, 485 (9th Cir. 1967) (quoting Note, *Standing to Sue For Treble Damages Under Section 4 of the Clayton Act*, 64 COLUM. L. REV. 570, 585 (1964)).

373. *See, e.g.*, Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956); *see also supra* notes 283-85 and accompanying text.

374. *See, e.g.*, Reibert v. Atlantic Richfield Co., 471 F.2d 727, 732-33 (10th Cir.), *cert. denied*, 411 U.S. 938 (1973); Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty & Sur. Cos., 382 F.2d 925, 929 (10th Cir. 1967).

375. A further extension of the historical argument is found in Mans v. Sunray DX Oil Co., 352 F. Supp. 1095 (N.D. Okla. 1971), in which the court associates the direct-injury requirement with original congressional intent. *Id.* at 1099. "Federal statutory law did not supplant the common law but incorporated it." *Id.* In this view the direct-injury rule had not been a judicial innovation; its removal would be. This seems to be the first appearance of the premise that Congress intended to incorporate a direct-injury standing requirement in the treble damage sections of the Sherman and Clayton Acts, later asserted in Associated General Contractors v. California State Council of Carpenters, 459 U.S. 519, 532-33 (1983). Another innovation in the conventional doctrinal history was the notion that the standing doctrines developed by the lower courts reflected an effort to implement the original intentions of Congress. *See infra* note 379 and accompanying text.

376. *See, e.g.*, Mulvey v. Samuel Goldwyn Prods., 433 F.2d 1073, 1076 (9th Cir. 1970), *cert. denied*, 402 U.S. 923 (1971).

377. *See, e.g.*, Hoopes v. Union Oil Co., 374 F.2d 480, 485 (9th Cir. 1967); Harman v. Valley Nat'l Bank, 339 F.2d 564, 567 (9th Cir. 1964); *see also* Radovich v. National Football League, 352 U.S. 445, 453-54 (1957) (in interpreting § 4 of the Clayton Act the courts "should not add requirements to burden the private plaintiff beyond what is specifically set forth by Congress"); Radiant Burners, Inc. v. Peoples Gas, Light & Coke Co., 364 U.S. 656, 659-60 (1961) (per curiam); Mandeville Island Farms, Inc. v. American Crystal Sugar Co., 334 U.S. 219, 236 (1948).

378. 405 U.S. 251 (1972).

"[t]he lower courts have been virtually unanimous in concluding that Congress did not intend the antitrust laws to provide a remedy in damages for all injuries that might conceivably be traced to an antitrust violation."[379]

The ten circuit court of appeals opinions cited by the Court for this proposition include both restrictive and expansive versions of direct-injury doctrine.[380] The selection seems calculated to avoid indicating a preference among any of the approaches represented.[381] Oblique as this reference to lower court case law is, however, the Court did not repeat its own previous warnings against artificial restrictions on standing. The Court's remark was generally viewed as approving some form of judicially created standing doctrine.[382]

The ordinary-meaning approach to construction of the treble-damage provision, which had governed antitrust standing analysis at the turn of the century, was regarded as a recent innovation in the 1970s because a false picture of the doctrinal history had taken hold.

---

379. *Id.* at 264 n.14. The lower courts promoting the direct-injury rule and its variants generally had not asserted that their versions of standing doctrine embodied the concrete intentions of the Congress of 1890 in enacting § 7. For example, of the ten circuit court of appeals decisions the court cited in support of this statement, *see id.*, none asserts this. Only three of this group make even conclusory references to congressional purpose relevant to standing. *See* Billy Baxter, Inc. v. Coca-Cola Co., 431 F.2d 183, 187 (2d Cir. 1970) (deterrent purpose of treble-damage remedy requires certainty of application), *cert. denied*, 401 U.S. 923 (1971); *Hoopes*, 374 F.2d at 485 (recognizing standing of plaintiff "serves the statute's purposes"); Sanitary Milk Producers v. Bergjans Farm Dairy, Inc., 368 F.2d 679, 689 (8th Cir. 1966) (plaintiff is "one which § 4 of the Clayton Act . . . is designed to protect"). Otherwise these three decisions, like the rest of the cited decisions and like the modern cases generally, concentrate on discussing how earlier courts had addressed standing issues. One of the cited decisions supports the direct-injury rule by invoking Congress's inaction in the face of judicial pronouncements, not Congress's original purpose. Nationwide Auto Appraiser Servs., Inc. v. Association of Casualty & Sur. Cos., 382 F.2d 925, 929 (10th Cir. 1967). The only prior case that had attributed a generally applicable standing restriction to congressional intent seems to have been *Mans*, 352 F. Supp. at 1099 (Sherman Antitrust Act "incorporated" direct-injury rule based on common law).

An interesting example of how far afield antitrust standing doctrine had strayed from principles of statutory construction is found in Malamud v. Sinclair Oil Corp., 521 F.2d 1142 (6th Cir. 1975). In applying administrative law principles to the question of antitrust standing, the court treated § 1 of the Sherman Antitrust Act rather than § 4 of the Clayton Act as the relevant statutory provision. *Id.* at 1151-52.

380. *Hawaii v. Standard Oil*, 405 U.S. at 264 n.14.

381. *Compare, e.g., Billy Baxter, Inc.*, 431 F.2d 183 *with Hoopes*, 374 F.2d 480.

382. *See, e.g., Malamud*, 521 F.2d 1142; *In re* Multidistrict Vehicle Air Pollution M.D.L. No. 31, 481 F.2d 122, 126 (9th Cir.), *cert. denied*, 414 U.S. 1045 (1973); *cf.* LAWRENCE A. SULLIVAN, HANDBOOK OF THE LAW OF ANTITRUST 774 n.23 (1977) ("This observation does not commit the Court to all or any part of the judicial gloss which shimmers in the standing area.").

The ordinary-meaning legacy had already been forgotten by the 1950s; it remained forgotten. Although courts of the 1950s adopted the direct-injury rule on grounds of policy, it eventually came to seem so natural and authoritative that the myth of its early origin acquired credibility. By the mid-1970s courts viewed direct-injury doctrine as the original interpretation of the pertinent statute. The ordinary-meaning approach was correspondingly spurned as an implausible novelty,[383] and it never reappeared in treble-damages standing jurisprudence.

IV. Looking Ahead: Doctrinal Implications

Section 7 of the Sherman Antitrust Act and its successor, section 4 of the Clayton Act, use the broadest terms in authorizing private treble-damage suits: any person injured in his business or property, by reason of any violation, may sue.[384] Standing doctrine, however, extensively restricts the range of persons permitted to use this remedy. The history of this development, investigated above, prompts a doctrinal question: is this judicial constriction of the statutory remedy legitimate?

The previous section argues that the debate over the legitimacy of standing doctrine in the 1960s and 1970s was critically affected by the inaccurate account of this doctrine's history as it was then accepted. This section will assess the contribution an improved understanding of that doctrinal history might make toward antitrust standing analysis today. This article shows that many points in the conventional history of antitrust standing doctrine are erroneous. Given the central part such erroneous historical representations play in legitimating current antitrust standing doctrine, today's doctrine deserves extensive reexamination.

While the Supreme Court's most recent decisions on antitrust standing doctrine prescribe a multipart balancing test, within that test directness of injury remains central.[385] The Court's reconciliation of restrictive standing doctrine with the broad statutory language rests on several related historical propositions in which the direct-injury rule also is central. Relying on the conventional history, the

---

383. *See In re* Multidistrict Vehicle Air Pollution M.D.L. No. 31, 481 F.2d 122, 125 & n.4 (9th Cir.), *rev'g* 52 F.R.D. 398, 401 (C.D. Cal. 1970), *cert. denied*, 414 U.S. 1045 (1973). Wilson v. Ringsby Truck Lines, Inc., 320 F. Supp. 699 (D. Colo. 1970), was disapproved in Reibert v. Atlantic Richfield Co., 471 F.2d 727, 732-33 (10th Cir.), *cert. denied*, 411 U.S. 938 (1973). *See supra* note 362 and accompanying text.

384. 15 U.S.C. § 15 (1973 & Supp. 1990).

385. *See, e.g.*, Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 540 (1983); Blue Shield v. McCready, 457 U.S. 465, 476-78 (1982).

Court states that in enacting the Sherman Antitrust Act, Congress intended for the courts to use standing restrictions comparable to the direct-injury rule;[386] that the courts, attempting to infer Congress's intent regarding section 7, always have used a restrictive approach to the treble-damage provisions, beginning with the direct-injury rule;[387] and that in enacting the Clayton Act, Congress intended to incorporate the direct-injury rule itself because the courts were already using it.[388]

This Article shows that the historical prominence of the direct-injury rule has been greatly exaggerated. The conventional premise that the direct-injury rule governed antitrust standing doctrine from the start is incorrect: from 1890 to 1914 only one published opinion contained any discernible reference to a generalized direct-injury requirement, and that reasoning was both secondary and ambiguous.[389] The ordinary-meaning approach, rather than the direct-injury rule, was the first dominant interpretive convention applied by the courts to questions of standing to seek antitrust treble damages.[390]

These facts are of enormous consequence. The Supreme Court's recent inference that the Congress of 1890 intended for the courts to use restrictions comparable to the direct-injury rule in interpreting section 7 of the Sherman Antitrust Act is not based on the legislative record, which contains no reference to any such intention.[391] Instead, it is inferred largely from the perception that the courts construing section 7 between 1890 and 1914 perceived such a congressional intention and implemented it.[392] Because this proposition is incorrect—those courts in fact perceived that Congress had intended the treble-damage provision to be read generally according to its ordinary meaning for purposes of standing analysis[393]—the case for attributing the direct-injury rule to the Congress of 1890 is fatally weakened.

A related point applies to section 4 of the Clayton Act, the source of today's treble-damages remedy.[394] The Court's recent assertion that Congress intended to adopt a generalized direct-injury rule when it enacted section 4 depends in turn on the proposition that the courts had adopted such a restriction between 1890 and 1914.[395] Again, because the courts did not adopt such a restriction,

386. *See Associated Gen. Contractors*, 459 U.S. at 532-34.
387. *See id.* at 530, 533-34.
388. *See id.* at 534; *see also supra* notes 9-12 and accompanying text.
389. *See supra* notes 118-31 and accompanying text.
390. *See supra* notes 68-74, 80-81 and accompanying text.
391. *See* Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 533 (1983).
392. *Id.* at 532-34.
393. *See supra* notes 68-74, 90-92 and accompanying text.
394. 15 U.S.C. § 15 (1973 & Supp. 1990).
395. *See Associated Gen. Contractors*, 459 U.S. at 534.

the argument that Congress intended to incorporate direct-injury standing principles in the Clayton Act is also virtually unsupported.[396] Especially when the legislative record offers no support for restrictions,[397] the early construction of section 7 is powerful evidence that a generally applicable direct-injury requirement should be rejected as an "artificial limitation" on the treble-damage remedy.[398]

Because the direct-injury concept of standing lacks any substantial basis in the legislation or case law of 1890-1914, any historical argument in favor of preserving a generally applicable directness concept in today's standing doctrine would have to rely on some notion that the direct-injury rule enjoyed long and continuous recognition beginning sometime after 1914. Such an argument would encounter serious obstacles in the historical record, however. The dominance of the unitary direct-injury rule is a relatively recent phenomenon. Between 1915 and 1950 a bare handful of decisions invoked a direct-injury rule; most courts avoided it.[399] Only in the 1950s, sixty years after the original enactment of the treble-damage remedy, did the direct-injury rule suddenly become dominant.[400] The direct-injury rule cannot be regarded as having become solidly established until after the final repudiation of the ordinary-meaning approach to the treble-damage provision in the 1970s.[401]

An argument that relies on the precedential weight of the direct-injury rule must also take into account the use by many of the modern direct-injury decisions of definitions of directness inconsistent with the concept of directness found in the few pre-1950 direct-injury decisions.[402] With respect to particular fact situations, few concrete standing restrictions have consistently enjoyed judicial consensus.[403] Even the well-known rules against stockholders, creditors, and officers seeking treble damages for losses caused by harm to a corporation have not always enjoyed unanimous assent among the courts.[404]

Congress made no counter-response after the rise of the direct-injury standing rule in the 1950s, and an argument might be made that congressional inaction implied ratification of the direct-injury rule. Congress's lack of response to the rise of the direct-injury rule in the 1950s, however, should not be viewed as tacit ratification when the courts promoting the new rule invoked a false doctrinal

---

396. *See id.* at 534.
397. *See supra* notes 10, 391 and accompanying text.
398. Blue Shield v. McCready, 457 U.S. 465, 472 (1982).
399. *See supra* notes 182-228 and accompanying text.
400. *See supra* notes 266-76 and accompanying text.
401. *See supra* notes 362, 376-84 and accompanying text.
402. *See supra* notes 274, 354 and accompanying text.
403. *See, e.g.*, Berger & Bernstein, *supra* note 5, at 820-24.
404. *See supra* notes 89, 331-33, 342 and accompanying text.

history to support it,[405] which implied that Congress already had acquiesced in the direct-injury rule for four decades,[406] and that no novel development in the antitrust standing case law had occurred for Congress to scrutinize.

Another aspect of the actual historical development of the direct-injury rule tends to deflate further this rule's authority as support for today's doctrine. Underlying the Supreme Court's recent invocation of the precedential value of direct-injury doctrine is the premise that the lower courts which fashioned this doctrine had done so in an effort to implement what Congress had intended.[407] This premise is incorrect. None of the decisions that invoked a unitary direct-injury rule before the 1950s ascribed this requirement to congressional intent.[408] In the 1950s a few direct-injury decisions invoked dictum from Supreme Court opinions regarding the purpose of other parts of the antitrust laws,[409] but most did not even claim to be doing that; none cited evidence from the legislative history directly supporting the direct-injury rule.[410] Only in the 1970s did wishful courts begin to attribute the direct-injury rule to Congressional intent,[411] out of apparent concern regarding the lack of connection between judicial doctrine and the statute on which it was based. To the extent any continuous judicial tradition based on the direct-injury rule is discernible since the 1950s, it represents not a settled conclusion to the question of what Congress intended, but only a settled conclusion that this was an appropriate scope for the treble-damage remedy.

The Court has declared that "in the absence of some articulable considerations of statutory policy suggesting a contrary conclusion in a particular factual setting," section 4 of the Clayton Act should be applied "in accordance with its plain language and its broad remedial and deterrent objectives."[412] The two policy objectives that apparently led courts to adopt a generalized direct-injury standing rule—limiting the cumulative size of damage awards assessed against particular defendants[413] and limiting the burdens imposed on the

---

405. *See* Karseal Corp. v. Richfield Oil Corp., 221 F.2d 358, 363 (9th Cir. 1955); Conference of Studio Unions v. Loew's Inc., 193 F.2d 51, 54 (9th Cir. 1951), *cert. denied*, 342 U.S. 919 (1952).

406. *See* Snow Crest Beverages, Inc. v. Recipe Foods, Inc., 147 F. Supp. 907, 909 (D. Mass. 1956).

407. *See* Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 533-34 (1983).

408. *See, e.g.*, United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574 (2d Cir. 1916); *see also supra* notes 182-89 and accompanying text.

409. *See supra* notes 280, 285 and accompanying text.

410. *See supra* note 283 and accompanying text.

411. *See supra* note 375 and accompanying text.

412. Blue Shield v. McCready, 457 U.S. 465, 473 (1982).

413. *See supra* notes 302-03 and accompanying text.

lower courts by private antitrust litigation[414]—are both flatly contrary to the language of section 4 of the Clayton Act.[415] Neither of these can be regarded as a statutory policy of that provision.

Today's antitrust standing jurisprudence is statutory construction only in the broadest sense: courts give the words of the treble-damage provision no real weight, relying on the premise that section 7 of the Sherman Antitrust Act, like section 1, "cannot mean what it says."[416] This statement would have baffled those who worked with section 7 between 1890 and 1914. No such argument was even proposed to contemporary courts, which regarded the language of the treble-damage section as concretely meaningful in its ordinary sense.[417] This ordinary-meaning approach to antitrust standing questions antedates the first assertion of a general direct-injury standing requirement.[418]

The early case law also permits us to better evaluate the Supreme Court's statement that the treble-damage section does not "provide a remedy in damages for all injuries that might conceivably be traced to an antitrust violation."[419] In the context of the early decisions, this proposition is inconsequential: it is not at all inconsistent with an ordinary-meaning approach to the treble-damage section. Two examples will suffice. First, the ordinary-meaning interpretive convention accommodated the common-law rule against individual stockholder damage actions for harm to the corporation.[420] Second, the rule of damages barring recovery for speculative damages meant that some "conceivable" theories of recovery would be rejected without reference to standing doctrine.[421] Even the courts that promoted the direct-injury doctrine in the 1950s never claimed that the treble-damage language could not mean what it seemed to mean; they invoked supposedly authoritative judicial usage.[422]

Once the premise that section 7 cannot mean what it says is set aside, little support exists for the Supreme Court's fundamental proposition that Congress intended to give the courts wide latitude

---

414. *See supra* notes 306-15 and accompanying text.

415. *See* 15 U.S.C. § 15 (1973 & Supp. 1990) ("That *any* person who shall be injured in his business or property by reason of *anything* forbidden in the antitrust laws may sue therefore . . . .") (emphasis added).

416. Associated Gen. Contractors v. California State Council of Carpenters, 459 U.S. 519, 531 (1983) (quoting National Soc'y of Professional Eng'rs v. United States, 435 U.S. 679, 687-88 (1978)).

417. *See supra* notes 68-74 and accompanying text.

418. *See supra* notes 118-19, 182-84 and accompanying text.

419. *Associated Gen. Contractors*, 459 U.S. at 534 (quoting Hawaii v. Standard Oil Co., 405 U.S. 251, 263 n.14 (1972)).

420. *See supra* notes 81-85 and accompanying text.

421. *See supra* notes 77-79, 93 and accompanying text.

422. *See supra* notes 257-59, 279-84 and accompanying text. Many decisions emphasized the gap between the import of the statutory language and the restrictive judicial doctrines. *See supra* notes 284, 373-74 and accompanying text.

in developing standing restrictions that would limit the scope of section 7.[423] The "naturally broad and inclusive meaning" of the treble-damage section[424] was also the original meaning: the basis of the dominant interpretive convention during the period between the Sherman Antitrust Act and the Clayton Act. This historical evidence provides powerful support for the proposition that standing doctrine should once again give the broad language of the treble-damage section presumptive weight.

CONCLUSION

The ability of courts to successfully alter legal rules may ultimately be constrained only by the need to present arguments that "achieve the necessary degree of plausibility" in the context of "the existing legal tradition."[425] This seems to explain the lack of judicial consensus in favor of the public-injury standing rule, for example, despite its attractions for courts increasingly burdened by antitrust litigation.[426] The degree of constraint exerted by the legal tradition at any given time, however, depends on the current state of knowledge regarding that tradition. The less guidance is provided by serious historical investigation, the weaker such constraint will be. In the case of antitrust standing doctrine the longstanding lack of curiosity regarding its history yielded quite a minimal constraint.[427] Not surprisingly, those courts aggressively promoting innovation in treble-damage standing doctrine have enjoyed relatively free rein in their pursuit of reasonable outcomes. Nor is it surprising that when well-meaning judges concoct a plausible history identifying desired norms in the past, this wishful thinking eventually becomes accepted as historical truth.

---

423. *See supra* notes 9-13 and accompanying text.

424. Blue Shield v. McCready, 457 U.S. 465, 473 (1982) (quoting Reiter v. Sonotone Corp., 442 U.S. 330, 338 (1979)).

425. WATSON, *supra* note 358, at 98.

426. *See supra* notes 317-28 and accompanying text.

427. *See supra* notes 363-71 and accompanying text.

This article is a chapter in LANGUAGE IN THE JUDICIAL PROCESS (Judith N. Levi & Anne Graffam Walker, editors) and is reprinted here by permission of Plenum Publishing Corp., 233 Spring Street, New York, NY 10013.

Chapter 11

ADEQUACY OF CIGARETTE PACKAGE WARNINGS: AN ANALYSIS OF THE ADEQUACY OF FEDERALLY MANDATED CIGARETTE PACKAGE WARNINGS

BETHANY K. DUMAS

1. INTRODUCTION[1]

The recent rise of interest in health warnings has coincided with federal legislation mandating new, rotating warnings on cigarette packages (Comprehensive Smoking Education Act of 1984); the warning requirement has also been extended to smokeless tobacco products, including snuff and chewing

[1] I want to express my gratitude to my editors, Judith N. Levi and Anne Graffam Walker, for the uncompromising rigor of their standards and for their tireless encouragement and assistance. Both made important substantive as well as stylistic contributions to this chapter. I am particularly grateful to Professor Levi for suggesting to me the role of space limitations in the wording of the cigarette package warnings. I also want to acknowledge the assistance of colleagues at the University of Tennessee in performing statistical analysis and planning and preparing displays of the data treated herein. Special thanks are due to Ralph G. O'Brien, of the Department of Statistics and the Computing Center, who prepared Figures 1–3; thanks are also due to Donald R. Ploch, of the Department of Sociology and Milton D. Broach and Michael O'Neil of the Computing Center. Finally, I want to express my gratitude to John Karnes, who provided prescription drug labels, and to all those persons who assisted in the research by responding to questions posed during Experiments 1, 2, and 3.

BETHANY K. DUMAS • Department of English, University of Tennessee, Knoxville, Tennessee 37996-0430.

tobacco (Comprehensive Smokeless Tobacco Health Education Act of 1986). The new warning requirements stem from the increased awareness on the part of the public health advocates of possible health risks (Blasi & Monaghan, 1986) and also from increased efforts by lobbying groups to further restrict both cigarette advertising (Abramson, 1986:1) and permissible smoking in public areas.

There has also been a new round of litigation in which cigarette smokers or survivors of deceased cigarette smokers have attempted to recover damages from cigarette manufacturers on the theory that addictive smoking of cigarettes is responsible for maladies and conditions such as (1) lung cancer, (2) emphysema, and (3) limb amputation resulting from cardiovascular problems. Most of the lawsuits against tobacco companies in recent years have been wrongful death cases in which plaintiffs have charged that tobacco manufacturers are responsible for deaths or illnesses resulting from these maladies and conditions.

From the point of view of some of plaintiffs' litigators, the current cases constitute Phase 2 of an ongoing cigarette litigation saga (Lee, 1986). Phase 1 occurred between 1954 and the late 1970s. Although the tobacco manufacturers lost no cases during Phase 1, cases brought during that period sometimes resulted in findings by juries that smoking cigarettes causes lung cancer, but that tobacco manufacturers could not have known that prior to the issuance of the Surgeon General's first report on smoking and lung cancer in 1964 (Office of Smoking and Health, U.S. Department of Health, Education and Welfare, *Smoking and Health: A Report of the Surgeon General* [1964]) and so were not liable. After a several-year lull in litigation, Phase 2 "got into full swing in published opinions in 1984" (Lee, 1986:21); Phase 2 cases differ from Phase 1 cases in part because of the existence of federally mandated warnings from 1965 to the present (Federal Cigarette Labeling and Advertising Act of 1965, Public Health Cigarette Smoking Act of 1969). Some litigators predict that Phase 3 will involve mass litigation and ultimate success for plaintiffs (Gidmark, 1986:8–9; Lee, 1986:23).

These factors—the general rise of interest in health warnings, the new federal legislation mandating warnings on tobacco products, and renewed litigation involving claims against tobacco product manufacturers—suggest the need for research into the adequacy of warnings on cigarette packages. The concern of this chapter is the adequacy of past and present federally mandated warnings on cigarette packages.

This chapter begins by identifying the legal issues implicated in cigarette warning litigation, after which it identifies some issues of adequacy from the point of view of linguistics and human factors analysis. It then discusses what is known about the structure and function of warnings and describes strategies for exploring how individuals interpret warnings in general, then warnings on cigarette packages in particular. The chapter concludes that there is strong evidence for the existence of objective criteria by means of which the

relative adequacy of warnings on cigarette packages can be assessed and that those objective criteria are for the most part not characteristic of present or past cigarette package warnings. Suggestions are offered for strengthening existing warnings, and additional research is proposed.

It should be understood by the reader at the outset that it is controversial whether written or graphic warnings ever constitute effective forms of safety information. The presence of written or graphic warnings on products seems mandated primarily by *common sense*, on the basis that explicit information will, in fact, be heeded by consumers (Lehto & Miller, 1988:5). However, some researchers have argued that warning labels play only a small role in the propensity of consumers to heed or disregard information found on them.[2]

It should also be understood that the warnings on cigarette packages have a different history and perform a different function from warnings on consumer products that must be used in order to accomplish specific tasks and also from warnings on pharmaceutical products whose prescribed use is beneficial to the consumer, though perhaps associated with certain health hazards or negative side effects. There is no reported long-term benign or therapeutic physical effect of the daily cigarette smoking habits that characterize most American smokers.[3] Thus the task of the tobacco industry in promoting its products is to sell, on the basis of short-term satisfaction, products with severe long-term negative consequences for both consumers and others. In drafting cigarette warnings, Congress has had to walk a tightrope between warning the public of health hazards associated with use of the product and avoiding angering the tobacco industry, a powerful economic force that is still heavily subsidized by our government.

The smoking of cigarettes is still legal in our society, yet there is strong

[2] One researcher (Breznitz, 1984) has explored the psychology of false alarms in early warning systems and has demonstrated that the effectiveness of warnings may be reduced by the *false alarm effect*, which operates to reduce (1) the credibility of the warnings and (2) the willingness of individuals to engage in protective behavior. On the basis of that research, Breznitz has suggested that consumers eventually ignore health and safety information because they engage in a continuous process of self-shaping based on false alarms: "Consider, for instance, the case of smoking. In spite of all the information to the contrary, one smokes a cigarette and nothing happens. One smokes another cigarette and still nothing happens. Thus, in the absence of any clear signals that may indicate the danger involved, these threats turn out subjectively to be false alarms" (p. 232). (The alert reader will note that Breznitz has used *threats* as a synonym for *warnings*. The similarities between *threats* and *warnings* are discussed later in this chapter.) (See also Weinstein *et al.*, 1978:63.)

One design engineer, Roger L. McCarthy, has had great success on the witness stand as a defendant's expert claiming that "the presence or absence of a warning absolutely makes no measurable effect on safety-related behavior, period." To the best of my knowledge, McCarthy has not substantiated that claim with empirical evidence (Deposition, 1987:156).

[3] The *psychological satisfaction* reported by such smokers is generally regarded by the medical profession as constituting "relief of the nicotine withdrawal syndrome by the self-administration of the drug in the form of tobacco" ("A Glossary of Tobacco Terms," 1985, citing Samfield, 1980:102, 151).

evidence that smoking is both addictive and deadly. In fact, there is mounting evidence that there is no safe use of cigarettes. So long, however, as cigarette smoking remains legal, it may be impossible to persuade consumers that smoking is deadly. At the very least, the decision to smoke cigarettes may be the result of such a complex rationalization process that there is no reasonable way to affect it. Alternatively, it may be that in the case of such a product there are more effective ways of providing safety information to consumers than by placing warning labels on packages. However, under present law, the requirement that cigarette packages carry warning labels is a major way of providing safety information to consumers. It is, therefore, highly appropriate to analyze these warnings in order to assess their adequacy and effectiveness.

1.1. The Legal Issues

Two legal issues are basic to all contemporary tobacco product cases: causality and failure to warn. In law, causality is what is known as a threshold issue: It must be proved before the issue of failure to warn can become important. Because causality is a matter for medical expertise, it will not be discussed here; it is worthy of note, however, that the issue of causality presents problems to medical researchers because of the complexity of the interrelationship between environmental factors and individual human heredity.

Failure to warn is also a potentially crucial legal issue. Manufacturers of dangerous products can be held liable for failure to warn consumers of the dangers of using such products and can become liable for enormous sums of money. In our society, we have three mechanisms for controlling dangerous products. Some products are controlled by decisions of federal agencies such as the Food and Drug Administration. Other products, like cigarettes, can be manufactured and sold only if they carry warnings that are mandated by Congress. Finally, there is the control provided through common law, which is case law evolved not from legislation but from judicial opinion based on precedents in our legal system.

An examination of case law demonstrates that there are three general guidelines for consumer product warnings:

1. Safety warnings must be displayed on products if such products would be *unreasonably dangerous* without such warnings (Restatement [Second] of Torts § 402A, Comment j [1965]).[4]

[4] This section of the Restatement also contains a widely-accepted definition of *unreasonably dangerous*, under which an article is held to be unreasonably dangerous only if it is "dangerous to an extent beyond that which would be contemplated by the ordinary consumer who purchases it, with the ordinary knowledge common to the community as to its characteristics." The section goes on to give examples of dangerous products whose characteristics are sufficiently well known that they are not deemed unreasonably dangerous. Tobacco is given as an example of such a product: "Good tobacco is not unreasonably dangerous merely because the effects of smoking may be harmful; but tobacco containing something like marijuana may be unreasonably dangerous."

2. Warnings should be directed to the ultimate users of the product or to any individuals who might be expected to come into contact with the product. (Restatement [Second] of Torts § 388, Comment a [1965].)
3. To be adequate, a warning should (1) catch the attention of a reasonably prudent person in the circumstances of its use, (2) be understandable, and (3) convey a fair indication of the nature and extent of the potential danger to the individual (*Ford Motor Co. v. Nowak*, 1982).

The requirement of adequacy is the vaguest of the guidelines. Some of the reasons the courts have found manufacturers in violation of the requirement of adequacy have been (1) failure to communicate the level of the danger (*Bituminous Casualty Corp. v. Black and Decker Corp.*, 1974), (2) failure to display the warning in a reasonable location (*Griggs v. Firestone Tire and Rubber*, 1975), (3) failure to give sufficient information about how to avoid the danger (*Wallinger v. Martin Stamping and Stove Co.*, 1968), and (4) failure to preserve the integrity of the warning by including statements that nullify its impact (*American Optical v. Weiderhamer*, 1980).

Because, in the case of cigarette package warnings, the precise wording is mandated by Congress, the question arises: Is mere literal compliance with the statutory requirements sufficient to constitute adequacy of warning? That is, has Congress preempted the common-law guideline requiring that warnings, to be adequate, must (1) catch the attention of a reasonably prudent person in the circumstances of its use, (2) be understandable, and (3) convey a fair indication of the nature and extent of the potential danger to the individual? In one legal interpretation of the required warnings (*Roysdon v. R. J. Reynolds Tobacco Company*, 1986, discussed later), such literal compliance with statutory law constitutes adequacy as a matter of law. The question is important both to lawyers, on the one hand, and forensic linguists and human factors analysts, on the other, because matters of law are decided by the court (in the person of the judge); only matters of fact are the province of the jury. Thus once a judge rules that the warnings are adequate as a matter of law—as some have so ruled—then that issue cannot go to the jury, and no expert testimony on the subject can be offered during trial. Given such a circumstance, a linguistic or human factors analysis of warning adequacy would be rendered moot.

If, however, a cigarette manufacturer's responsibility is held to go beyond mere literal compliance with the statutory law, then the legal issue does become a potentially interesting one to researchers in linguistics and human factors. To forensic linguists, the issue presents the opportunity to study and render an opinion on a number of syntactic, semantic, and pragmatic factors. To human factors researchers, the issue presents the opportunity to study and render an opinion on such factors as size, placement, and visibility of warnings. These factors, each of which plays a part in the effectiveness of warnings as perceived by the public, will be featured in the discussion to follow.

The research reported in this chapter was conducted in the context of a trial in which plaintiff's lawyers anticipated that the adequacy of warnings on cigarette packages would be at issue. In the course of conducting that research, I reached some conclusions about the relative adequacy of past and present cigarette package warnings. The reader should note that the research was carried out during a relatively brief period of time and that it should thus be regarded as a pilot study. In my final section, I describe additional research, which I plan to carry out in the future.

1.2. The Adequacy Issues

In order to address the issue of adequacy of warning, we must first of all know when someone has been warned. The first question for the researcher is then: What constitutes a warning? In this chapter, I shall be concerned only with written statements whose wording explicitly identifies them as warnings. Further, I shall approach the question primarily from the point of view of the recipient, rather than the sender. Whichever point of view is adopted, it is generally acknowledged by philosophers (e.g., Searle, 1969), linguists (e.g., Fraser, 1975), and others (e.g., Lehto & Miller, 1986) who have studied warnings that the formalization of the semantic and syntactic rules for warnings is difficult.[5]

Formalizing the rules for warnings is difficult for two reasons. First, no discipline recognizes a clear, unambiguous definition of *warning*. Cross-discipline uncertainty about the precise meaning of the term is even greater than that within discrete disciplines (e.g., linguistics). Second, warnings may be either direct or indirect and either literal or nonliteral. That is, many warnings are highly context-dependent, and their interpretation may depend upon lesser or greater amounts of inferencing.

The uncertainty of the definition of the term "warning" derives largely from (1) its being confused with such activities as *instructing, persuading,* and *advising* and (2) its being associated with a wide variety of functions, different ones of which are stressed by different sectors of society (Lehto & Miller, 1986:15).

Dictionaries recognize the uncertainty of definition, partly by acknowledging that the activity of *warning* includes one or more of, but not necessarily all of, such activities as *informing, telling, counseling,* and the like. They also typically recognize close similarities among such words as "warn," "forewarn," "admonish," and "caution." *Warn* is generally stipulated as being the most comprehensive of such words; the common element of meaning among them is generally regarded as *giving notice of either actual danger or the possibility of danger.*

[5] See Kreckel (1981a, b) for the suggestion that there may be differing cultural notions about what it is to be warned.

An examination of definitions of the term *warning* in various disciplines underscores the uncertainty reflected in dictionary definitions. This uncertainty will be discussed later with respect to linguistics and philosophy. Only in human factors engineering textbooks is great consistency of emphasis displayed with respect to warnings. There, *alerting to a necessity for action* (including negative action, e.g., refraining from smoking) seems to be the major component of a warning (Lehto & Miller, 1986:14, citing De Greene, 1970:313; McCormick, 1970:189; Murrell, 1969:156, 208; Robinson, 1977).

Consumer products often arrive with "[a] wide variety of literature . . . expected to perform several functions that are commonly viewed as being warning-related" (Lehto & Miller, 1986:15). Different segments of society stress different functions of warnings, depending upon the nature of their stake in the product under examination; thus, there are varying perspectives on warnings. "These perspectives can simplistically be divided into the views of society as a whole and the views of the directly affected parties, which include [,among others,] manufacturers . . . and consumers" (Lehto & Miller, 1986:15). The general societal view of warnings is that they function primarily to reduce accidents and health damage by informing people of risks associated with the use of consumer products (Weinstein, Twerski, Piehler, & Donaher, 1978).

From the perspective of manufacturers, warnings may function largely as a defense against litigation (Lehto & Miller, 1986:16). Using a warning for such a purpose may lead to its actually having an *antisafety* function, that is, warnings may be used as a replacement for careful design (Schwartz & Driver, 1983).

But what about the perspective of the consumer? It is from this perspective that issues of adequacy arise. In the absence of a clear and unambiguous definition of warning, some researchers have accepted as a working definition that of Searle, a language philosopher who described warnings as *statements about future events or states which are not in the hearer's best interest, and which are uttered in situations in which it is not obvious to both the hearer and the speaker that the event will occur or that the state will transpire* (1969, p. 67). I used that working definition as a starting point for the research described here.

In positing the illocutionary act rules for warnings, Searle suggested that warnings are either categorical or hypothetical (1969:67). His definition (given before) was of categorical warnings, which are, he suggested, "like advising, rather than requesting" (1969:67). Hypothetical warnings are explicitly predictive: They follow IF–THEN logical structure and are of the type "If X, then Y." Searle suggested that most warnings are probably in this "hypothetical" category. An example is "If you do not do X, then Y will occur" (1969:67).

Searle has not, so far as I know, offered any evidence for these generalizations. The research reported later explored the distinction between categorical and hypothetical warnings, partly to test subjects' sensitivity to the difference, partly to test the adequacy of the classification.

Fraser (1975) pointed out that it is difficult to isolate warnings. From a strictly linguistic point of view, it is difficult to distinguish them from threats and promises. Fraser analyzed warning acts in a speech act framework and concluded that threats are a special type of warning, one wherein "the speaker takes on the responsibility for bringing about the disadvantageous action" (Fraser, 1975:173), whereas a promise is a specification of a future act in which "the speaker intends the utterance to count as the undertaking of an obligation to carry out the action" (Fraser, 1975:175). The close similarity between warnings and promises is illustrated by the frequent occurrence in informal conversation of the joking rejoinder, "Is that a threat or a promise?"

According to Fraser, then, warnings are like both threats and promises in that, as noted, they refer exclusively to future actions. Warnings are like threats and unlike promises in that they refer to undesirable future events. Warnings are unlike both promises and threats in that the future action or nonaction referred to is in the control of the hearer, not the speaker. Warnings are also unlike threats and promises in that they are never absolute, being always conditional or contingent, even though categorical warnings are only implicatively conditional or contingent. Though both promises and threats can be contingent, they can also be absolute as in "I promise I'll never do that again" and "I'll get you for this."

Further, as noted, warnings, like other specific illocutionary acts, can be indirect as well as direct, nonliteral as well as literal. Fraser (1975) has pointed out that warnings can be either verbal or nonverbal and that no particular words are necessary to warn. A raised eyebrow may constitute a stern warning at a dinner party. The words "Look out!" in a physically dangerous situation can certainly constitute a warning. And, again as noted, there are differing notions of the functions of warnings.

I have said that I am concerned here only with written warnings identified by explicit wording as being warnings. With respect to warning labels, it is important to note that the *warning label*, as distinct from a *warning* per se, has functions other than that of introducing a sincere warning. One function appears to be a mere lip service compliance with the letter of the law. Note, for instance, the following two current rotating *warnings* on cigarette packages:

1. SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risks to Your Health.
2. SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide.

The first example can be read as a sincere warning only if one allows for a negative inference—"If I don't quit smoking now, I continue to sustain serious risks to my health." As phrased, the statement is actually more like an explicit promise or a forecast because explicitly positive results are mentioned.

The second example, on the other hand, mentions no results of any kind

and thus appears to be merely informational. A consumer who is unaware of the effect of carbon monoxide on the human body could not, therefore, be relied upon to read the second example as a warning. In both cases, each of which will be discussed at greater length later, the intent seems to be merely to comply on a surface level with the statutory requirement. (The first example offers a further problem, that of the uncertainty triggered by the word "now." An addicted consumer may well be prompted to rationalize away the message of this warning by asking ask one of these questions: *Would my quitting smoking last year not have greatly reduced serious risks to my health? Would my quitting smoking next year not greatly reduce serious risks to my health?* This is surely nit-picking, but it is nit-picking that was engaged in by some of the subjects I interviewed in the research described later.)

Another function of the *warning label* is merely to direct attention to an item. Consider the legend on the cover of a Chinese restaurant menu in Knoxville, Tennessee:

WARNING!
Chinese Food Lovers
Eating Any Selection From The Enclosed MENU
Can Be Dangerously Habit Forming!

Clearly this tongue-in-cheek announcement is made not to warn away but to entice customers to try the food because no self-respecting restaurant owner would seriously suggest that his or her food is clinically addictive.

Difficult as it is to identify warnings and to distinguish them from warning labels, it is a great deal more difficult to know what constitutes an adequate or effective warning. One of the processes humans appear to be engaging in when they read warnings is that of risk assessment. Risk assessment is an extremely complex process, partly because of the inherent complexity of the process whereby consumers process and act on any information. On the basis of the research reported, it seems clear that some of the linguistic factors that affect the extent to which consumers heed warnings include those of syntax, semantics, and pragmatics. Others appear to be such variables as size, placement, and visibility of warnings. Without claiming to address the full scope of the adequacy of warning issue, in this chapter I will undertake to identify some objective criteria of warning messages reported effective or ranked high on an effectiveness scale; I will then assess past and present federally mandated cigarette package warnings with respect to whether they are characterized by those objective criteria.[6]

[6] Since completing the empirical research described in this chapter, I have become familiar with the books by Lehto and Miller (1986) and Miller and Lehto (1987) cited before. *Volume I, Warnings: Fundamentals, Design, and Evaluation Methodologies* (Lehto & Miller, 1986), focuses on safety information labels on products from the perspective of human factors and product design engineering. It contains a chapter, "Definitions and Modeling Techniques" and three chapters

1.3. Genesis of the Study

The research reported here had its genesis in a request by a plaintiff's lawyer in Knoxville, Tennessee, that I conduct research on the issue of the adequacy of past and present federally mandated cigarette package warnings. Between late October and early December of 1985, I carried out an empirical investigation in preparation for the submission of a report in the case styled as *Roysdon v. R. J. Reynolds Tobacco Company* (1986), a case in which plaintiff Roysdon claimed that his leg amputation (necessitated by cardiovascular disease) was the result of his having smoked cigarettes manufactured by R. J. Reynolds Tobacco Company for most of his life.

In planning my research, I focused primarily on the six warnings that either have been required to be printed on cigarette packages since 1965 or are required now. The plaintiff in the case had smoked since before 1965, so he had smoked when there were no required warnings, continued to smoke during the tenure of the 1965 and 1970 warnings, and on into the early months of the new rotating warnings. Statutes of limitations precluded his suing under the no-warning period or the time of the 1965 warning (1965–1970). However, I felt that comprehensive testing of all these warnings plus some fabricated ones was necessary for comparison of the effect of various linguistic and extralinguistic factors in warnings. The six federally mandated warnings I planned to test are given below in order of promulgation; 3 through 6 are the rotating warnings currently in force:

1. Caution: Cigarette Smoking May Be Hazardous to Your Health (1965)
2. Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous to Your Health (1970)
3. SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema and May Complicate Pregnancy (1985)
4. SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risks to Your Health (1985)

---

on the effectiveness of warnings. Lehto and Miller use a knowledge-based approach (in the sense used in artificial intelligence) to human performance and task analysis and advocate strongly the need to address risk rationally, rather than on the basis of *common sense*. Though the volume does not include the perspectives of linguistics and philosophy, it is an extremely important book for anyone seriously interested in issues involving consumer product warning formulation and adequacy. Volume II is an *Annotated Bibliography with Topical Index* (Miller & Lehto, 1987). A third volume, tentatively titled *Formalized Design Standards and Design Methods for Compliance* (Lehto, Miller, & Clark) has been announced for publication in late 1988 or early 1989. It will contain, among other things, a summary of the warnings required by the federal government; descriptions of warning/labeling standards as functions of products; identification of products required to have warnings; and the specific nature of the warnings that must be designed for the products. It is also to contain an illustrative expert system designed to assist in the design and layout of physical warning labels.

5. SURGEON GENERAL'S WARNING: Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight (1985)
6. SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide (1985)

The general legal issue in the context of which I was asked to conduct my research was whether past or present cigarette package warning labels are adequate to warn consumers of the possibly negative effects of cigarette smoking. Specifically, I was asked to give an opinion as to whether the 1970 warning is adequate to warn consumers of the potentially negative cardiovascular effects of cigarette smoking.

Shortly before trial, however, Judge Thomas G. Hull of the Eastern District of Tennessee, the presiding judge, rendered my testimony inadmissible by ruling that, as a matter of law, the federally mandated warnings on cigarette packages are adequate. This ruling had the effect of removing the issue from the jury's consideration of it. Further, at the end of the plaintiff's presentation of his evidence, Judge Hull dismissed the suit, ruling that the plaintiff had made no case. Judge Hull gave two reasons for doing so: (1) The federal statute on cigarette package labeling had preempted state common-law actions based on alleged inadequacies, and (2) common knowledge about tobacco was such that cigarettes are not *unreasonably dangerous*.[7] The plaintiff appealed, but the Sixth Circuit upheld the ruling of the trial court. No attempt will be made by the plaintiff to appeal to the Supreme Court of the United States, so defendant R. J. Reynolds won. Although my prepared testimony therefore did not play an important role at the trial, it did provide the impetus for further study of the issues in dispute.

As suggested, my overall goal in preparing for trial had been to conduct research designed to shed light on this general question: Are past or present federally mandated cigarette package warnings, particularly the 1970 warning, adequate to warn smokers of the harmful effects of smoking, particularly the potentially negative cardiovascular effects of cigarette smoking? In order to explore this general question, I addressed the following specific questions:

1. Can warnings be identified by the presence of specific semantic, lexical, syntactic, or other characteristics? If so, what are they?
2. Can warnings be meaningfully classified? If so, on what bases? Do consumers respond to or evaluate warnings on such bases?
3. Is it possible to state categorically how warnings are perceived by consumers? If so, how are warnings perceived?
4. Do warnings differ by degree? Do consumers classify warnings as strong versus weak?

[7] Judge Hull noted that Tennessee tort law has been held to incorporate the comment defining *unreasonably dangerous* in the Restatement (Second) of Torts, § 402A, discussed in footnote 4.

5. Are the federally mandated cigarette package warnings strong or weak? If they are weak, how could they be made strong?
6. Do the statements at issue here constitute warnings? If so, precisely what do the statements appear to warn against?
7. Are the warnings at issue adequate to warn consumers of possible negative effects of cigarette smoking, particularly with regard to the potentially negative cardiovascular effects of cigarette smoking? If they are not, could they have been?

In addition, I wanted to investigate the extent to which there is uniformity in how individuals understand warnings. Should there be a great uniformity in the effect of some of the factors I had identified earlier (and which I discuss later), such evidence might eventually suggest a theoretical model for the formalization of linguistic rules concerning adequate warnings. At a minimum, such evidence would suggest a general framework for analysis of the adequacy of many consumer products.

1.4. Initial Analysis of Warnings on Cigarette Packages

In my initial analysis of the warnings, I familiarized myself with the legislative history of the federal promulgation of the warning label requirements on cigarette packages and with the role of the Federal Trade Commission (FTC) in the history of that promulgation. After that, I examined some federally mandated warning labels used on prescription drugs. Finally, I analyzed the six cigarette warnings in order to identify potential problems with content and readability and to formulate hypotheses for research.

Legislative history is made up of legislative background documents and events, including committee reports, hearings, and floor debates; it is used by courts when they are required to determine the legislative intent of a particular statute. I read the legislative history of the warning label requirements on cigarette packages in order to isolate the factors identified as important by the drafters of these warnings. From it, I learned that the warnings that have actually been required have always been significantly weaker (as determined by criteria I report later in this chapter) than those initially proposed. For instance, legislation mandating use of the original (1965) warning "Caution: Cigarette Smoking May Be Hazardous to Your Health") actually had the effect of preempting a proposed FTC Trade Regulation Rule that would have required all cigarette packages and advertisements to disclose clearly and prominently that "Cigarette Smoking is dangerous to health and may cause death from cancer and other diseases." (The Trade Regulation Rule for the Prevention of Unfair or Deceptive Advertising and Labeling of Cigarettes in Relation to the Health Hazards of Smoking, 29 Fed. Reg. §§ 8324, 8325, 8356 [July 2, 1964] ["The 1964 Cigarette Rule"].) Had Congress not preempted the requirement of The 1964 Cigarette Rule, the very first federally mandated

warning would have mentioned specific negative consequences of smoking cigarettes, namely specific diseases, and would have specified that smoking *is dangerous*, not merely that it *may be hazardous*.

Later, in 1969, shortly before the end of a congressionally mandated moratorium on requiring cigarette manufacturers to reveal tar and nicotine contents of cigarettes, the FTC proposed a modified version of The 1964 Cigarette Rule; had that modified version of the original proposal been adopted, it would have required all cigarette packages and advertisements to carry this message: "Warning: Cigarette Smoking Is Dangerous to Health and May Cause Death From Cancer, Coronary Heart Disease, Chronic Bronchitis, Pulmonary Emphysema, and Other Diseases." (34 Fed. Reg. § 7919 [1969]) Congress, however, amended the message on cigarette packages to read: "Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous To Your Health." (15 U.S.C. §§ 1331 *et seq.*, 1970).

Since then, the four rotational warnings in current use have been adopted. Again, the impetus for change seems to have come from the FTC. A May 1981 FTC Staff Report on the Cigarette Advertising Investigation ("Staff Report") sets out the reasons why that agency thought the 1970 warning was ineffective. The first factors, identified on the basis of *common sense*, were (1) overexposure (the warning was "worn out"), (2) lack of novelty (it contained no new information), (3) abstract and general nature of the wording, and (4) lack of personal relevance of the warning. Also, the unchanging size and shape of the 1970 warning were felt to contribute to its ineffectiveness. Later market research surveys reported by the FTC suggested additionally that (1) if warnings were to be effective, they should be short (one idea per warning), simple, and direct, and (2) disease-specific warnings, that is, those listing specific diseases as possible consequences of smoking, are far more effective than non-disease-specific warnings.

The proposal to use a rotational warning system evolved partly as a way to address the four problems cited before. The FTC recommended that the rotational warnings should be selected in accord with four criteria: (1) medical accuracy, (2) demonstrable filling of a gap in consumer knowledge about health hazards, (3) intelligibility, and (4) ability to "prompt consumers to think about the health hazards of smoking" (Staff Report pp. 5–33). Sample warnings prepared by the FTC meet all those criteria. Representative ones include the following:

1.	WARNING:	Smoking causes death from cancer, heart attacks and lung disease.
2.	CARBON MONOXIDE:	Cigarette smoke contains carbon monoxide and other poison gases.
3.	WARNING:	Smoking may be addictive.
4.	LIGHT SMOKING:	Even a few cigarettes a day are dangerous.

Again, we find the same pattern of FTC-proposed warnings mentioning specific negative consequences of smoking, followed by Congressionally promulgated warnings mentioning fewer or weaker specific negative consequences of smoking. The four rotating warnings currently required are a good deal weaker and certainly less comprehensive than the first ones proposed by the FTC.[8]

After having familiarized myself with the legislative background, I next examined some federally mandated warning labels used on prescription drugs. In order to do so, I obtained a large sampling from a Knoxville, Tennessee, pharmacist. In examining them, I noticed that the pharmaceutical labels appeared to observe the criteria listed here very well. Most of the warnings are brief; additionally, the labels appear to be medically accurate; and they certainly did fill gaps in my own knowledge about health hazards. They were generally intelligible, and, as a consumer, I felt that they generally would prompt me to think about the health hazards of any prescription drugs associated with them. The most striking aspect of the prescription warnings was the use of graphic symbols (such as automobiles and outlines of faces) and color contrast. On the basis of this examination, I decided that it was obviously possible for warning label designers to comply strongly with those four FTC-proposed criteria: The resulting warnings could be medically accurate; they could fill gaps in consumer knowledge about health hazards; they could be intelligible; and they could "prompt consumers to think about the health hazards of smoking."

The FTC seemed to expect that the introduction of rotational warnings would correct the prior problems of overexposure and lack of novelty, while recognizing that the problems of the abstract nature of the prior warnings, and the perceived lack of personal relevance, would have to be corrected by a judicious choice of wording for each of the rotating warnings. The choice of wording would in addition have to conform to the following criteria:

1. Medical accuracy
2. Demonstrable filling of a gap in consumer knowledge about health hazards
3. Intelligibility

[8] This chapter is not an appropriate forum for discussing the process whereby legislation is promulgated by our Congress. I would, however, like to note that one important difference between Congress and federal regulatory agencies like the FTC is that Congress is the target of systematic lobbying by private interest groups, often with interests at odds with those of the average consumer and often with enormous sums of money at their disposal. One such group is the Tobacco Institute, the lobbying arm of the tobacco industry. The wording of cigarette package warnings in this country has always been the result of a compromise between Congressional proposals and successful lobbying efforts of the Tobacco Institute and other private groups.

4. Ability to direct consumers' attention to health consequences of smoking (possibly by keeping warnings short (one idea per warning), simple, and direct, and by creating disease-specific warnings)
5. Specificity and concreteness
6. Personal relevance
7. Variation in format

Criterion 1 is an issue for medical research; it will not be discussed further here. Discussed here are criteria 2 through 7, particularly criteria 3 through 7. At issue are both the actual importance of all these criteria and also the extent to which current warnings comply with them.

1.5. Hypotheses

Using what I had learned from completing the steps discussed, I analyzed the six cigarette warnings for the purpose of identifying potential problems with content, readability, and format. I considered (1) vocabulary, (2) syntax, (3) warning type (using Searle's distinction between categorical and hypothetical warnings), and (4) identity of addressee. I also identified some potential problems with content (e.g., some labels not constituting warnings *per se*; others being too narrowly targeted to be personally relevant to most smokers; still others containing modal auxiliaries like "may" and "can") and with readability (e.g., use of technical vocabulary and unusual syntax). After that, I identified additional potential problems based on extralinguistic factors such as size, placement, and visibility of warnings. At this stage, I formulated the following set of general hypotheses with respect to the warnings under consideration:

1. When asked, respondents rank-order warnings with great consistency.
2. Specific, identifiable factors are characteristic of those warnings selected as stronger.
3. Other things being equal, respondents rank hypothetical warnings as stronger than categorical warnings.
4. The modal auxiliaries "may" and "can" weaken warnings; further, the presence of one of these hedging modals anywhere in a warning has the effect of being taken to be generally applicable to all clauses in the warning.
5. The presence of either modal auxiliary "may" or "can" in a warning in a rotational series has the effect of being read into all the warnings in the series.
6. Disease-specific warnings are perceived as stronger than non-disease-specific warnings.

7. Unusual syntax, as in the *double -ing* construction (*Quitting Smoking Now . . .*), weakens warnings.
8. The presence of a nonwarning (*Quitting Smoking Now . . .*) instead of a warning in a statement labeled as a warning weakens the statement's potential for alerting to health risk.
9. The presence of a nonwarning in a rotational series of warnings has the effect of reducing the effectiveness of all the warnings in the series.
10. Technical or semitechnical terms, which may be misunderstood by some consumers, render warnings less effective. (One of my graduate students reported that some women in birth classes attended by lower socioeconomic families in Oak Ridge, Tennessee, read *Fetal Injury* as *Fatal Injury*. Other women in those classes thought that *Low Birth Weight* was a desirable result of smoking. It was unclear whether that was because they interpreted the statement to mean that *their* weight would be lower at birth or that having a baby weighing less would be desirable. It is difficult to propose a practical solution to the problem illustrated here because it results partly from the very restrictive space limitations imposed upon the cigarette package warnings. This issue is discussed in detail later.)
11. Warnings that are hard to read because of type size or color contrast are ignored.
12. Warnings that are placed so that they can be ignored (i.e., consumers opening a package of cigarettes or removing a cigarette from a package avoid looking at such a warning).

2. EXPERIMENT 1: CATEGORIZATION, RANK ORDERING, AND TRANSLATION OF WARNINGS

In Experiment 1, I explored primarily general hypotheses 1 and 3: (1) When asked, respondents rank order warnings with great consistency and (2) other things being equal, respondents rank hypothetical warnings as stronger than categorical warnings. Hypothesis 1 stemmed from my suspicion that consumers process public information in similar ways; Hypothesis 3 stemmed from my conviction that the IF–THEN logical structure of hypothetical warnings would trigger consistent classification; also, I expected that hypothetical warnings would be recognized as warnings *per se* more easily than categorical warnings and would for that reason also trigger more consistent classification. I also gathered some data with respect to all the other hypotheses (except for 10); because these data are not quantifiable, I defer further mention of them until the discussion section.

In order to test Hypothesis 1, I needed to know first of all whether there

was a consistent pattern of rank ordering of warnings; then I needed to know whether the same characteristics showed up consistently in those warnings ranked strongest. In order to test Hypothesis 3, I needed to know initially whether the distinction between categorical and hypothetical warnings was susceptible to empirical testing, that is, whether respondents could distinguish between them consistently on the basis on brief instruction. I also wanted to know how respondents would translate warnings if asked to rewrite them in their own words.

I thus designed a questionnaire, "Reactions to Warnings," which asked respondents to perform four tasks: (1) identify warnings as either categorical or hypothetical, on the basis of brief instruction, (2) select the stronger warning of several pairs of warnings, (3) rank order a set of 10 warnings as to strength, and (4) translate the 1970 warning into their own words.

I administered the questionnaire to 27 students enrolled in two beginning linguistics classes (one undergraduate class, one class with both undergraduates and graduates) taught by me at the University of Tennessee during the fall quarter of 1985. I will discuss each task and responses to it separately.[9]

2.1. Task I: Categorization of Warnings

2.1.1. Methods

The 16 warnings that were provided to subjects in this experiment are shown here in the order of presentation in the experiment:

1. [DRAWING OF A SKULL AND CROSSBONES] POISON
2. Take heed, sweet soul.
3. You have a right to remain silent. If you choose not to remain silent anything you say or write can and will be used against you in a court of law. You have a right to consult a lawyer before any questioning and you have a right to have the attorney present with you during any questioning. You not only have a right to consult with a lawyer before any questioning but, if you lack the financial ability to retain a lawyer, a lawyer will be appointed to represent you before any questioning. If you choose not to remain silent and do not wish to consult with a lawyer or have the lawyer present, you still have the right to remain silent and you have a right to consult with a lawyer at anytime during the questioning.
4. Let me tell you something straight. When you go and snitch to anyone that we had anything to do with this, you'll find a snitch tattoo on your forehead.

[9] Readers interested in replicating the experiments described in this chapter may obtain copies of the warning labels tested by requesting them from the author (see also Figure 3).

5. [DRAWING OF AN AUTOMOBILE] Caution This Drug alone or with alcohol may impair your ability to drive
6. Komsing Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy
7. Warning: The Surgeon General Has Determined That Komsing Is Hazardous to Your Health
8. This is the final warning. I have acted as a gentleman should, have given you ample time to consider my demands before an unfortunate incident occurs. You have twenty-four hours to introduce a bill in the Congress of the United States of America to return to me, as the rightful heir to James Smithson, the Smithsonian Institution and its belongings. Time has run out, sirs.
9. [DRAWING OF A PROFILE OF A FACE WITH AN OPEN MOUTH] FOR ORAL USE ONLY
10. SURGEON GENERAL'S WARNING: Quitting Komsing Now Greatly Reduces Serious Risks to Your Health.
11. The Surgeon General Has Determined That Komsing Is Dangerous to Your Health
12. [DRAWING OF A PROFILE OF A FACE WITH A HALF-CLOSED EYE] may cause DROWSINESS USE CARE when operating a car or other Dangerous machinery
13. SURGEON GENERAL'S WARNING: Etteragic Komse Contains Carbon Monoxide
14. Komsing by Pregnant Women May Result in Fetal Injury, Premature Birth and Low Birth Weight
15. Warning: Komsing is Dangerous to Health and May Cause Death from Cancer and Other Diseases
16. Etteragic Komsing May be Hazardous to your Health

All the examples provided for the respondents are "real" warnings—that is, they were collected from actual sources in the world, rather than being fabricated. Eight of the warnings are thinly disguised cigarette package warnings (warning No. 15 was once proposed but never implemented). Thus the base element "smok-" of the word "smoking" appears in reverse (as "koms-," hence "komsing") and the word "cigarette" is spelled in reverse ("etteragic") in every instance. (The disguising was done only in Task I, in which I was interested solely in whether respondents could distinguish the two types of warnings. In the remainder of the questionnaire the cigarette warning labels are reproduced without disguise.)

Four of the other warnings (Nos. 1, 5, 9, 12) are pharmaceutical labels in present or past use. Warning 3 is the familiar *Miranda* warning, the use of which is mandated in many arrest scenarios. Two warnings, 2 and 8, come from literature, the first from the final act of Shakespeare's *Othello*; I obtained it by asking the senior Shakespearean in the English Department at the

University of Tennessee to provide me with a statement from literature that he was positive was a warning, though it did not look like one. The latter, 8, is from Margaret Truman's novel, *Death at the Smithsonian*. The final warning, 4, I extracted from a surreptitiously tape-recorded conversation that I had analyzed for a Knoxville defense attorney in a solicitation/conspiracy to commit murder case.

The reader may be curious to know why I tested an assortment of both oral and written warnings at this stage when I have made it clear that my concern is with written warnings. In the earliest stages of my research, I was seeking common factors among effective warnings, regardless of whether they were oral or written. Thus in preparing the written questionnaire, I made no attempt to exclude oral warnings. Some of the warnings on the questionnaire are highly problematic, particularly (as I discuss later) the very long *Miranda* warning, with its embedded hypothetical warning and the highly decontextualized Shakespearean quotation.

In Task I, respondents were asked to categorize 16 warnings as either categorical or hypothetical. I provided the respondents with a brief categorization of the two types of warnings and then asked them to categorize the 16 warnings by labeling them with a "C" or an "H." The students were asked to categorize the warnings on the basis of the following instructions:[10]

> The English language allows two types of warnings. One type consists of what are called categorical warnings. Categorical warnings are generally felt to fulfill the function of advising, not requesting. Such warnings inform hearers or readers that certain results will follow certain modes of behavior, but the warnings do not attempt to get a given individual to modify his or her behavior. An example, tongue-in-cheek, is a statement on a menu that says, "Eating Any Selection From The Enclosed MENU Can Be Dangerously Habit Forming!"
>
> The other type consists of what are called hypothetical warnings. Hypothetical warnings are phrased in such a way that they fulfill the function of requesting. The basic logical structure for a hypothetical warning is *If X, then Y*, though the *if* and *then* may be implicit, rather than explicit. Thus an example might be *Give me your money or I'll shoot*, a statement which is generally regarded as a demand that the individual spoken to turn over money. (It is also a special type of warning, a threat.) There is occasionally a question as to which type of warning a particular statement is, so I would like to get your reactions to some warnings.

2.1.2. Data Analysis

In Experiment 1, I tested primarily the general hypotheses stating that (1) respondents rank order warnings with great consistency and (2) other things

[10] I now realize that the instructions for Task I were not as clear as they could have been. It is possible that some of the responses were equally divided in the way they were because of the way the instructions were worded, that is, students may have been sufficiently confused that they simply guessed at the answers. In future replications of this experiment, I shall word the instructions more carefully and also select clearer examples.

being equal, respondents rank hypothetical warnings as stronger than categorical warnings. In order to test them, I formulated the following sets of narrow hypotheses about the stimuli in the test instrument:

H1. Hypothetical warnings are recognized as such by a large majority of respondents.
H2. Categorical warnings are recognized as such by a smaller percentage of respondents.
H3. It is possible to identify characteristics that favor the identification of categorical warnings as categorical.
H4. Directions and incomplete statements are judged to be categorical by about half the respondents, as hypothetical by about half the respondents (because they will be guessing).
H5. Other things being equal, hypothetical warnings are judged to be stronger than categorical warnings.

Of the 16 warnings, I classified 2 (3, 4) as unambiguously hypothetical, one (8) less certainly so, and 10 (5, 6, 7, 9, 10, 11, 13, 14, 15, and 16) as categorical. Of the remainder, 1 (1) is a statement so incomplete that the respondent must supply information in order to be able to categorize it, and 2 (2, 12) are directions rather than warnings.

The responses, summarized in Figure 1, strongly confirm narrow hypotheses H1, H2, H3, and H4. As expected (according to Hypothesis H1), two hypothetical warnings, 4 and 8, were strongly recognized as such. Also as expected (according to Hypothesis H2), there was less agreement about the categorical warnings. Also as expected (according to Hypothesis H4), incomplete warnings and directions were interpreted about equally as categorical and as hypothetical. For example, the incomplete warning in 1 (*POISON*) was

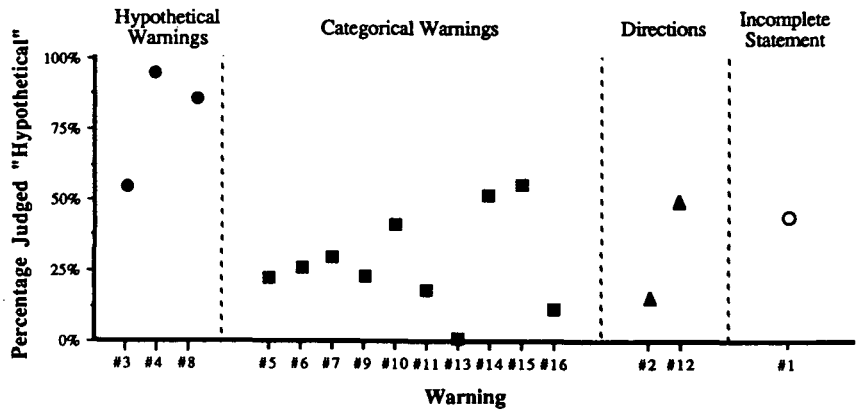


FIGURE 1. Percentage of subjects judging warning to be "hypothetical."

identified as categorical by 15 respondents, as hypothetical by 12. Similarly, the informational statement in 13 ("SURGEON GENERAL'S WARNING: Etteragic Komse Contains Carbon Monoxide") was identified as categorical by 13 respondents, as hypothetical by 13.

Narrow Hypothesis H5 was not tested directly here; as far as it is concerned, Experiment 1 was a phase for the gathering of information for use in later experiments. I will offer some observations about them in the summary section.

2.1.3. Discussion

Respondents showed some sensitivity to the difference between hypothetical and categorical warnings. With relatively brief instruction, they were highly consistent in their identifications. Only one respondent failed to respond to a hypothetical warning, whereas five respondents failed to respond to other types of warnings.

It was of considerable interest to me that the familiar *Miranda* warning was misidentified as categorical by close to half the respondents. I had originally expected it to be classified as a hypothetical warning by many subjects (and half of them did so) because of the embedded IF–THEN proposition, "If you choose not to remain silent, anything you say or write can and will be used against you in a court of law."

I think the results can be explained by these two reasons. First, the entire statement is a syntactically complex and illocutionarily quite complex text; it would really be quite surprising had subjects interpreted a lengthy and complex example in the same way that they identified one-sentence, one-clause examples. Further, the text of the *Miranda* warning is often truncated in films and television shows to the first 14 or so words, which inform suspects of rights: "You have a right to remain silent. If you choose not to remain silent. . . ." In fact, the responses to the *Miranda* warning illustrate the important nature of such factors as length, syntactic complexity, and illocutionary force.

I now think of the *Miranda* warning as a "hidden hypothetical (HH)"; by calling it that I do not intend to suggest that it constitutes a cognitively significant subcategory; rather, I intend the term to suggest a failure of intent. One intent of the *Miranda* warning was to put suspects on notice that their future rights could be jeopardized by how they chose to conduct themselves at the time of arrest; a failure to recognize the *Miranda* warning as a hypothetical warning may mean that the warning fails to achieve this purpose.

Also, I note with interest that the directive statement in warning 2 ("Take heed, sweet soul.") was identified as categorical by 22 respondents, as hypothetical by only 4. The consistency may be due to respondents' response to the high degree of inferencing required by the passage and its extreme decontex-

tualization. The line comes from the last act of Shakespeare's *Othello* and, in the context of the scene, clearly constitutes a warning with an implied consequence, for with it Othello continues his menacing questioning of Desdemona about the nature of her relationship with Cassio. However, I think it is possible that a modern equivalent statement, perhaps "Watch out, babe," would test differently, even out of context, simply because less inferencing would be required. I think that we probably more easily infer an implied consequence to "watch out" than we do to "take heed."

On the basis of responses to Task I, it appears that the following factors may favor the identification of warnings as categorical: (1) incompleteness, (2) decontextualization, and (3) the use of abstract and general terms as opposed to concrete and specific terms. Thus if it appears that hypothetical warnings are more effective than categorical warnings, it will be important to formulate them as complete statements, to avoid decontextualization, and to avoid couching them in abstract and general terms as opposed to concrete and specific ones. On the other hand, evidence will be offered below to the effect that the single word *POISON* is a very strong warning. Because it is arguably incomplete by almost any test, this issue needs further study.

2.2. Tasks II and III: Rank Ordering of Warnings

2.2.1. Methods

Uniformity of ranking and relative strength of warnings were tested further in Tasks II and III of the written questionnaire. In Task II, respondents were asked to select the stronger member in each of six pairs of warnings. In Task III respondents were asked to rank order ten warnings. Both Task II and Task III were described briefly in the final sentence of the general instructions for the entire questionnaire; immediately preceding Task II, the instruction to circle the stronger warning was given. The pairs are given below:

I. A. Cigarette Smoking May Be Hazardous to your Health
   B. The Surgeon General has Determined that Smoking Is Dangerous to Your Health

II. A. [Use whichever warning you found stronger in 1 and write its letter—A or B—here.]
   B. Smoking Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy

III. A. Smoking Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy
   B. [DRAWING OF A PROFILE OF A FACE WITH A HALF-CLOSED EYE] may cause DROWSINESS USE CARE when operating a car or other dangerous machinery

IV. A. Warning: Smoking Is Dangerous to Health and May Cause Death from Cancer and Other Diseases
B. [DRAWING OF A PROFILE OF A FACE WITH AN OPEN MOUTH] FOR ORAL USE ONLY

V. A. Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight
B. [DRAWING OF AN AUTOMOBILE] Caution This Drug alone or with alcohol may impair your ability to drive

VI. A. SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide
B. [DRAWING OF A SKULL AND CROSSBONES] POISON

2.2.2. Data Analysis of Task II

The students' responses to Task II, summarized in Figure 2, confirm general Hypothesis 1 that individuals rank order warnings with great consistency. On pairs I, II, and VI, almost all subjects (26/27, 24/24, and 26/26, respectively) judged warning B to be stronger; on pair III, close to 90% (23/26) judged warning A to be stronger. For pairs IV and V, subjects judged the statements to be more similar: Approximately 60% (16/26 in both cases) said warning A was stronger.

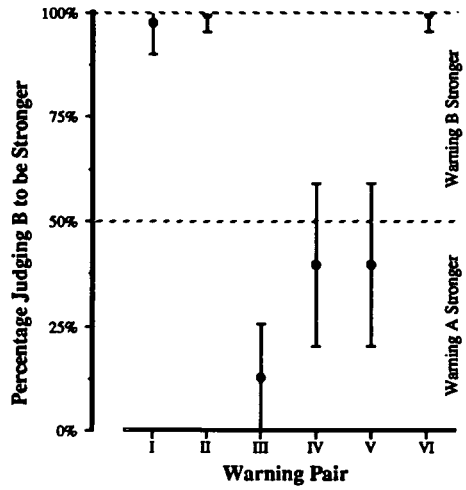


FIGURE 2. Percentage of subjects judging warning B to be stronger than warning A. Error bars represent approximately 95% confidence intervals for the population percentage. These intervals were obtained using %B ± 2[s.e.(%B)], where %B is the sample percentage based on $N = 26$ subjects and s.e. (%B) is estimated using the larger of SQRT[(%B)(100 − %B)/N] or SQRT[(99)(1)/N].

By their 26:1 choice, it is clear that respondents regard the 1970 cigarette package warning in I.B. as stronger than the original 1965 *Caution*, shown in I.A. It is at least equally clear (by a 24:0 margin) that the new disease-specific warning shown as II.B. is considered stronger than the 1970 warning. And, what is not surprising is that respondents chose the powerful *POISON* over a "warning" about carbon monoxide by a 26:0 margin. It is interesting that a disease-specific warning that says that "Smoking Is Dangerous to Health and May Cause Death from Cancer and Other Diseases" was rated weaker than one that said "For ORAL Use Only." Comments from respondents elsewhere on the questionnaire suggest that capital letters and drawings strengthen the impact of a warning. Respondents' comments also suggest that these two hypotheses would be well worth further investigation: (1) that listing a series of unpleasant consequences strengthens a warning and (2) that inclusion of modal auxiliaries that hedge an assertion, such as *may* or *could*, weaken warnings. Another possibility is that nonparallel structure (as in IV.A.) weakens a potentially strong warning because the weaker of two clauses may "contaminate" the stronger.

2.2.3. Data Analysis of Task III

In Task III, respondents were asked to rank order 10 warnings as to relative strength:

1. [DRAWING OF A PROFILE OF A FACE WITH AN OPEN MOUTH] FOR ORAL USE ONLY
2. Take heed, sweet soul.
3. You have a right to remain silent. If you choose not to remain silent anything you say or write can and will be used against you in a court of law. You have a right to consult a lawyer before any questioning, and you have a right to have the attorney present with you during any questioning. You not only have a right to consult with a lawyer before any questioning but, if you lack the financial ability to retain a lawyer, a lawyer will be appointed to represent you before any questioning. If you choose not to remain silent and do not wish to consult with a lawyer or have the lawyer present, you still have the right to remain silent and you have a right to consult with a lawyer at anytime during the questioning.
4. Let me tell you something straight. When you go and snitch to anyone that we had anything to do with this you'll find a snitch tattoo on your forehead.
5. [SKULL AND CROSSBONES] POISON
6. The Surgeon General has Determined that Smoking Is Hazardous to Your Health
7. This is the final warning. I have acted as a gentleman should, have

given you ample time to consider my demands before an unfortunate incident occurs. You have twenty-four hours to introduce a bill in the Congress of the United States of America to return to me, as the rightful heir to James Smithson, the Smithsonian Institution and its belongings. Time has run out, sirs.

8. The Surgeon General has Determined that Smoking is Dangerous to Your Health
9. Warning: Smoking Is Dangerous to Health and May Cause Death from Cancer and Other Diseases
10. Cigarette Smoking May be Hazardous to your health.

The responses, reflected in Figure 3, further address the issues of uniformity of ranking and relative strength of warnings and tend to confirm specific hypothesis one about consistency of ranking.

Though there was a wide distribution of rankings across the respondents, a pattern of uniformity emerged from an examination of the mean ranks and modes of the 10 warnings in Task III.[11] The uniformity of these rankings suggested that, although not everyone evaluates the strength of the warnings the same, there is enough of a tendency or pattern that such a methodology could be fruitfully used in a large-scale survey to determine more about how people respond to warnings. Thus, warning 5 (POISON [drawing of a skull and crossbones]) is rated strongest by all measures; it has a mean rank of 1.95 and a mode of 1. Warning 9 (the only disease-specific warning in the set) is rated second strongest with a mean rank of 3.79 and a mode of 2. Close behind it is warning 4 (a clear hypothetical type), rated third strongest with a mean rank of 4.37 and modes of 2 and 3. Warning 3 (*Miranda*) is rated fourth strongest with a mean rank of 5.2 and a mode of 3. Competing for fifth and sixth places are warnings 7 (the 1970 Surgeon General's Warning) and 6 ("This is the final warning"). Warning 7 has a mean rank of 5.33 and modes of 5 and 6, whereas warning 6 has a mean rank of 5.37 and a mode of 5. Warning 1 is rated eighth in strength with a mean rank of 6.08 and a mode of 6. Warnings 8 and 10 have mean ranks of 5.70 and 7.87, and modes of 6 and 7, respectively. Predictably, warning 2 ("Take heed, sweet soul") was rated weakest with a mean rank of 9.6 and a mode of 10.

2.3. Task IV: Translation of a Warning

2.3.1. Methods

Task IV was administered to respondents as a way to obtain some information about how they understood cigarette warnings. I thought it would be

[11] The *mean rank* or *average* is a measure of the frequency of distribution of responses; it is calculated by summing (adding) the individual rankings and then dividing by the total number of responses. The *mode* is the most frequently obtained score in the data.

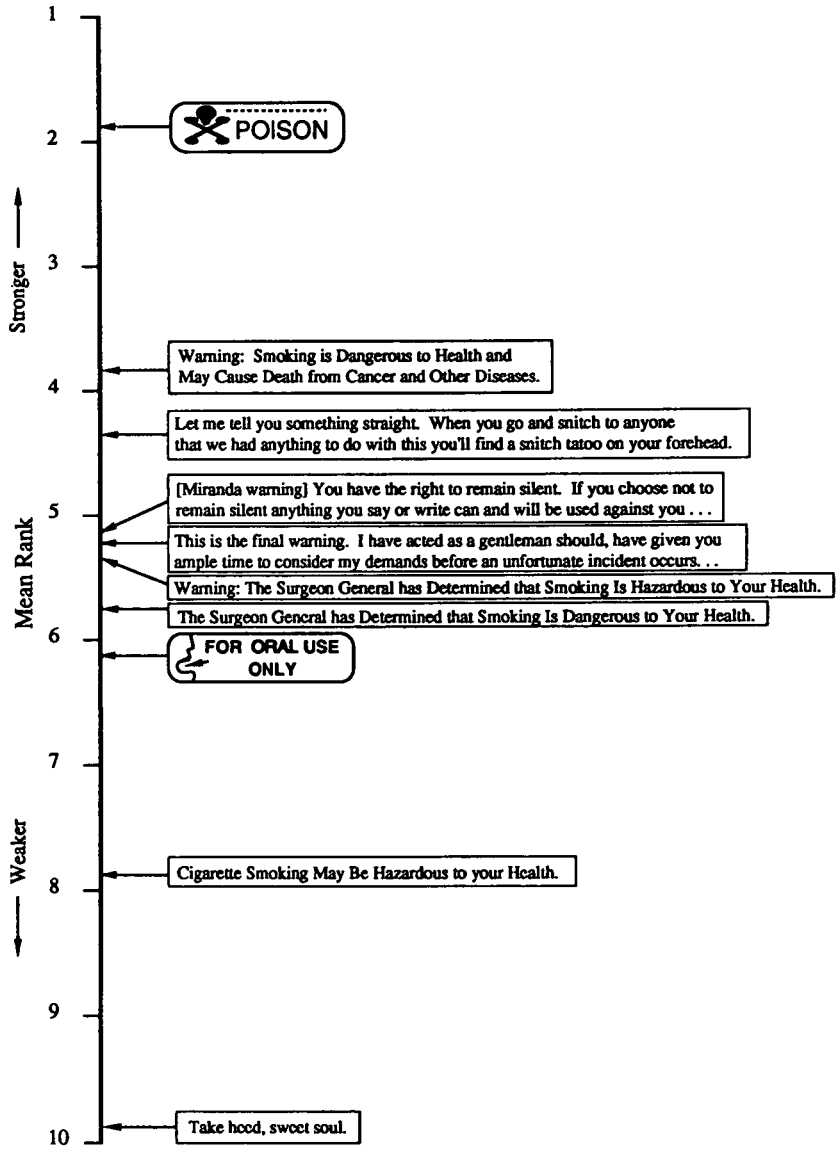


FIGURE 3. Rank-ordering of warnings in Task III of Experiment 1.

helpful to know what additions, deletions, and substitutions would be made, as well as what rearrangements of parts might occur as respondents paraphrased the words of the 1970 warning. In this task, respondents were given these instructions: "Translate the following warning into less formal language. Use your own words but try to reflect the exact meaning of the original warning." They were then given the text of the 1970 warning: "Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous to Your Health."

2.3.2. Data Analysis

The respondents' paraphrases of the crucial embedded clause, "Cigarette Smoking Is Dangerous to Your Health," are displayed in Table 1. The complete text of the warning consists of three parts: (1) an initial signal, *Warning:*; (2) identification of the authority whose decision is being announced, *The Surgeon General Has Determined (that)*; and (3) the crucial finding itself, *Cigarette Smoking Is Dangerous to Your Health*. In analyzing the paraphrases provided by respondents, I was most interested in the third part, which constitutes the actual warning. Table 1 shows the range of paraphrases supplied by subjects for this crucial section (followed by rewordings of the first two parts as well).

Note that although the main verb of the original warning clause is "is"—an unqualified, affirmative, present-tense form of the verb "be," 14 out of 27 respondents altered that verb in such a way as to significantly change its meaning. Many did this by adding a modal auxiliary such as "may" or "can." Some modals, of course, like "shall" and "will," serve merely to indicate simple futurity. Others, however, particularly "may," "might," "can," and "could," add notions of doubt to the predictive function of the verb *be*. Thus paraphrasing the words *is dangerous* as *may be dangerous* represents a significant semantic change in the verb phrase.

Other paraphrases added a "do" or "have" form. Also 13 changed the verb from an inflected form to the uninflected *be* or to one of the following lexical verbs: "make," "harm," "cause," "happen" (to your health), "die," "damage," "prove," "kill," or "attribute" (for "contribute?").

2.3.3. Discussion

What are the possible explanations for these changes, particularly the addition of those modal auxiliaries that inject doubt into the verb phrase? One explanation is that all cigarette package warnings are weakened by the fact that some of them contain qualifying language, sometimes in the form of the modal auxiliaries "may/might" and "can/could" (according to general hypothesis 4). The new disease-specific rotating warning, for instance, concludes with the clause "And May Complicate Pregnancy." The modal auxiliary "may"

TABLE 1. Written Paraphrases of 1970 Warning in Task IV of Experiment 1[a]

Subject	Auxiliaries (modals in caps)	Verb	Object or complement
cigarette smoking	MAY	make	you sick
smoking	WILL	be	harmful
smoking cigarettes	WILL	harm	the smoker
smoking		is	dangerous to your health
people who smoke		are	likely to get very sick
they [cigarettes]		cause	you to become sick or to get a disease
smoking cigarettes		is	dangerous
smoking		is	dangerous
something	COULD		happen to your health to seriously damage it
smoking		to be	bad for you
you	CAN	die	
smoking cigarettes	CAN and WILL	damage	your health
cigarette smoking		is	dangerous to your health
you	MAY	harm	your health
smoking		to cause	lung cancer
[you]	do not	smoke	
cigarette smoking	has been	found	to be dangerous to your health
cigarette smoking	MAY	be	very unhealthy
cigarette smoking	CAN	kill	
cigarette smoking	CAN	kill	you
the ill effects of cigarette smoking	have	prove to be not only unhealthy, but fatal	
smoking		is	unhealthy
smoking	MAY	attribute [*sic*] to poor health, sickness or disease	

[a]Markers, qualifiers, authorities, and main clause verbs included by respondents but omitted in the display are summarized here:

Markers: warning:, watch out, attention!, but be forewarned, please
Qualifiers: according to, this is, studies by
Authorities: surgeon general, the medical research, the surgeon general, surgeon general's, an authority, research, based on extensive studies
Verbs: believes, advise, has determined, has found out, declared, prove, shows, has proved, has been proven, warn

also appears in another new rotating warning: "Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Rate." These rotating warnings had been on packages in the market for several months at the time this task was administered, but packages displaying the old 1970 warning were also still available in the market. It is possible that respondents were reading into all warnings they were familiar with the qualifying language found on other warnings in current use.

It is also possible, of course, that the test instrument, the written questionnaire, was the source of contamination because it contained various examples of warnings containing qualifiers, including modal auxiliaries.

However, because the warnings on the written questionnaire mirror the fact that the new rotating warnings contain such qualifiers, it does not seem likely that the written questionnaire is the sole source of contamination.

It is also quite possible that the specific circumstances of the administration of the questionnaire (its being administered in a classroom situation) presented difficulties. Though I know of no literature on the subject, it is, for instance, possible that university students use modal auxiliaries in classroom paraphrase tasks as politeness markers.

2.4. General Discussion of Experiment 1

In general, narrow hypotheses H1, H2, and H3 were supported by the data. The evidence was that the distinction between categorical and hypothetical warnings is or can be made psychologically real, that respondents were consistent in selecting one of a pair of warnings as being stronger, and that they rank-ordered warnings consistently as to strength. The specific generalizations that emerge from close examination of the pilot data are these:

1. With the exception of conventional warning labels such as the word *POISON*, respondents identified hypothetical warnings as hypothetical more often than they recognized categorical warnings as categorical.
2. Respondents perceived hypothetical warnings as stronger than categorical warnings.
3. Extralinguistic factors, such as larger type size, use of drawings, and capital letters, appeared to strengthen warnings.
4. Finally, the presence of a qualified warning in a series of rotational warnings (i.e., the series that appeared on cigarette packages at the time this research was conducted) appeared to have a contagious weakening effect in that a warning that did not contain a qualifying modal auxiliary was paraphrased by close to a third of the respondents as containing such a modal auxiliary.

3. EXPERIMENT 2: IN-DEPTH INTERVIEWS

3.1. Methods

My next step began with a heuristic phase during which I developed additional research strategies. During that phase, in order to decide what strategies to use for testing further, I purchased cigarettes, conversed with cigarette vendors in a variety of marketplaces, and observed cigarette smokers in those marketplaces opening just-purchased packages of cigarettes.

I also decided to test fabricated warnings. That decision and my need to include in my testing materials a cigarette package labeled with the original (1965) *Caution* label led me to a decision to use in my next experiments only cigarette packages that were red and white in color. I reached that decision after noting that many packages, including those of generic brands, are red and white in color. I used a red typewriter ribbon to type all the warning labels I fabricated except for those involving the word *poison* and a drawing of a skull and crossbones. For those, I obtained printed labels from the pharmacist mentioned earlier. Those printed labels were also all red and white.

The tentative conclusions reached in Experiment 1, as well as the need to test additional hypotheses, formed the basis for a second research strategy, that of conducting increasingly complex interviews in which I asked interviewees to give me their reactions to different warnings. I began by using cigarette packages labeled with five different warnings, the 1970 *Warning* and the four current rotational warnings. At a midstage, I used cigarette packages containing seven different warnings, the five described before and also the original *Caution* and a fabricated disease-specific "Surgeon General's Warning" whose final part read: "Smoking Is Addictive and Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy." In the final interviews, I used a total of nine warnings, all those described before plus two additional ones bearing the word *POISON* between two drawings of a skull and crossbones. In warning 8, the POISON warning was on the side of the cigarette package, as is mandated by Congress. In warning 9, the POISON warning was again on the side of the package; but in addition, a smaller POISON label appeared on top of the package, where it would normally be opened. The full set of warnings explored in these interviews is shown below:

1. Caution: Cigarette Smoking May Be Hazardous to Your Health
2. Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous to Your Health.
3. SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide.
4. SURGEON GENERAL'S WARNING: Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight.
5. SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risk to Your Health.
6. SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy.
7. SURGEON GENERAL'S WARNING: Smoking is Addictive and Causes Lung Cancer, Heart Disease, Emphysema and May Complicate Pregnancy.
8. [SKULL AND CROSSBONES] POISON [SKULL AND CROSSBONES]

9. [SKULL AND CROSSBONES] POISON [SKULL AND CROSSBONES] [The package on which this warning was displayed also displayed a smaller, but otherwise identical warning on top, where it would normally be opened.]

My hypothesis was that warnings 2, 6, 7, 8, and 9 would be perceived as serious warnings, with 8 and 9 being the strongest and 7 the next strongest. I also hypothesized that the next strongest one would be 6. Behind it would be 2. I felt the others (1, 3, 4, 5) would be perceived as weak or ineffective, and perhaps even meaningless.

At this stage I was more interested in collecting open-ended information than I was in attempting to collect either the kind or amount of information that would be amenable to statistical analysis. I therefore conducted these interviews conversationally, jotting down comments of interviewees and noting the ordering they produced as they examined and physically rearranged the cigarette packages I showed them.

I conducted 24 of these comparative interviews, 6 at each stage. In these interviews, I showed the interviewees the cigarette packages one at a time and asked them to tell me what they noticed about the warnings. The precise questions I used are given below, in order of presentation:

1. [Show pack 2] What exactly does this warning mean to you?
2. [Show pack 1] Does this mean the same thing? The last pack said *Caution* and this one says *Warning*. Are those the same things? If not, how are they different?
3. [Show pack 3] Does this mean the same thing?
4. [Show pack 4] How about this one?
5. [Show pack 5] What does this one mean?
6. [Show pack 6] How about this one?
7. [Show pack 7] If you saw this warning on a pack of cigarettes, would it mean more or less to you—or something different—than the last one? How is it different?
8. [Show pack 8] What would you think this meant?
9. [Show pack 9] Please notice that this pack is exactly like the last one except that it has a small warning on the top also. Is the meaning of the warning the same if it's up there?

3.2. Data Analysis

The interviewees made responses to the individual questions, then often, as noted, also physically arranged the packages into various groupings, often describing to me the basis for the categorization. The categories were typically these: Very Specific, Very General, Very Exclusive (i.e., applies to only some

smokers), and Meaningless (sometimes Silly). Overall, the warnings were classified as shown next:

Effective, alarming	Any variant of a warning containing the word *POISON*
Very specific	SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy
	SURGEON GENERAL'S WARNING: Smoking Is Addictive and Causes Lung Cancer, Heart Disease, Emphysema, and May Complicate Pregnancy
Very general	Warning: The Surgeon General Has Determined That Cigarette Smoking Is Dangerous to Your Health
	SURGEON GENERAL'S WARNING: Quitting Smoking Now Greatly Reduces Serious Risks to Your Health
Applies to few	SURGEON GENERAL'S WARNING: Smoking by Pregnant Women May Result in Fetal Injury, Premature Birth, and Low Birth Weight
Silly, meaningless	Caution: Cigarette Smoking May Be Hazardous to Your Health
	SURGEON GENERAL'S WARNING: Cigarette Smoking May Be Hazardous to Your Health
	SURGEON GENERAL'S WARNING: Cigarette Smoke Contains Carbon Monoxide

3.3. Discussion

The comments by interviewees strongly supported the thrust of both my general and task-specific hypotheses:

1. Comments by respondents suggested that the modal auxiliary "may" tends to weaken warnings.
2. Disease-specific warnings are perceived as stronger than non-disease-specific warnings.
3. Unusual syntax, as in the *double -ing* construction (e.g., *Quitting Smoking Now*), renders warnings less effective. (Alternatively, the lack of specification of positive health effects may weaken this particular warning. However, respondents' remarks about this warning label—typically something like "it sounds funny"—suggest to me that the syntax plays a role.)
4. Incomplete warnings (e.g., "Cigarette Smoke Contains Carbon Monoxide") are sometimes perceived as silly.
5. Some respondents (who were not pregnant females) commented that they did not feel addressed by the warning addressed to pregnant

women and, further, that reading that warning caused them to take other warnings less seriously. This suggests that the inclusion of limited-target warnings in a rotational series may have the effect of weakening the overall effect of the series.

6. The conventional warning label *POISON* has great power. (For example, one young male smoker, when shown the package with *POISON* on the top said, "That scares the hell out of me!") It is unclear whether the power derives from familiarity or from some inherent force of the word *POISON* in combination with the drawing of a skull and crossbones.
7. Consumers expect printed warnings on consumer products to be conspicuous with respect to such factors as type size, color contrast, and position, i.e., consumers expect that statements that otherwise constitute warnings will not be nullified or rendered ineffective by factors rendering them difficult to read.[12]

As noted, I did not ask interviewees to rank-order the warnings in this experiment. I cannot, therefore, report a strict rank ordering here.

4. EXPERIMENT 3: RAPID, ANONYMOUS INTERVIEWS

4.1. Methods

Experiment 3, like Task IV of Experiment 1, addressed the issue of how respondents understand particular warnings. Unlike Task IV of Experiment 1, it sought spoken commentary rather than written paraphrase. In it, I obtained additional information by conducting rapid anonymous interviews. Such interviews are conducted by eliciting brief responses from unidentified strangers in a public setting (Labov, 1966, 1972). In the interviews, I asked interviewees to read one of three warnings on a cigarette package (the 1965 *Caution*, the 1970 Surgeon General's Warning, or the new disease-specific

[12] Factors such as type size, length of warnings, distance between viewers and ads, and movement between viewers and ads are reported to play a role in the readability of the Surgeon General's warning in billboards and taxicab ads. Davis and Kendrick (1989) concluded that the Surgeon General's warning is unreadable by consumers in its current form in most billboard and taxicab ads. A similar study conducted in Australia (Cullingford, Da Cruz, Webb, Shean, & Jamrozik, 1988) analyzed the apparent size of letters, color contrast between letters and their background, and the obliqueness of angle between the warning and the horizontal message and concluded that cigarette warnings on billboards are "a minor and, in many cases, illegible component of billboard advertisements for cigarettes in both comparative and absolute terms" (p. 338). See also Fischer, Richards, Berman, and Krugman (1989) for evidence, based on the market research methods of eye tracking, aided recall, and masked recognition, that the Congressionally mandated warnings are ineffective health messages insofar as adolescents are concerned.

warning) and then tell me in their own words exactly what it meant. I interviewed 47 persons (24 females, 23 males; 18 smokers, 26 nonsmokers) on a Sunday afternoon in West Town Mall, a large shopping center in Knoxville, Tennessee, using a clearly visible tape recorder to record their responses.

4.2. Data Analysis and Discussion

Few respondents limited themselves to translation of the warning they were shown. Many respondents commented on adequacy. However, the only question put to the respondents was a request that they translate the warning into their own words and tell me exactly what it meant. Respondents' translations and comments varied widely. Although most respondents repeated words in the warnings or paraphrased them in virtually identical language, a few just said that the warning meant "exactly what it says." Others claimed not to read the warnings. One respondent offered the opinion that the new warning addressed to pregnant women is severely limited in its effect and that a label with the word *POISON* on it might deter. Another offered the opinion that the word "may" means that smoking has not been proven to be hazardous to health.

There were no strong differences between smokers and nonsmokers or between those who were shown the 1965 *Caution* and those who were shown the 1970 *Warning*. Intuitively, I think that these data present some evidence that people process general (particularly non-consequence-specific) warnings very individually, in terms of such factors as their own values and risk-taking behavior. I intend to address these issues in a later stage of research (see "Directions for Future Research"). Both the cigarette smoking and the cigarette package warning issues are highly charged emotionally. Many of the persons I interviewed wanted to know which side I was on. One particularly rabid individual, a female smoker, used the occasion of our interview to tell me how much she hated television public announcements about the dangers of smoking. She said she leaves the room when one comes on. It is difficult to think of a warning label that she would regard as usefully informative.

5. SUMMARY AND CONCLUSIONS

5.1. Legal Issues

This chapter has been concerned with the linguistic and extralinguistic adequacy of written warnings on cigarette packages. I used a variety of strategies to address the questions introduced at the outset of my discussion, the answers to which are given later. I then developed a set of general hypotheses to guide me in my research. Each of those hypotheses is also

discussed briefly. First, however, I would like to review briefly some legal senses of *adequacy* and to summarize the evidence I gathered about the nature, function, and structure of warnings.

For purposes of this discussion, I will use the word "adequate" in its legal sense, insofar as its legal sense differs from its everyday sense. In law, it means "[s]ufficient; commensurate; equally efficient; equal to what is required; suitable to the case or occasion; satisfactory" or "[e]qual to some given occasion of work" (*Black's Law Dictionary* 36 (5th ed. 1979), citing *Nissen v. Miller* (1940)).

According to dicta (that is, statements that do not actually decide the case and hence have no precedential value) of the Supreme Court of the United States, "adequate" is a relative term; that is, adequacy must be assessed in the light of the facts and circumstances. In *Mullane v. Central Hanover Bank & Trust Co.* (1950), the Supreme Court decided that trying to give notice to interested parties of a pending disposition of a court case by publishing a notice in a newspaper was inadequate when there was available a way of contacting the parties directly, as from a mailing list. The Supreme Court then recognized the relative nature of adequacy when it said that notice by publication in a newspaper might be approved "as a customary substitute in another class of cases where it [was] not reasonably possible or practicable to give *more adequate warning* [emphasis added]" (p. 317).

Earlier I noted that where warning language is prescribed by Congress, it is possible that the courts may go in two directions on the issue of adequacy. In one legal interpretation of the required warnings, literal compliance with statutory law constitutes adequacy as a matter of law; cases have been and may well continue to be considered on that basis. It is also possible that a cigarette manufacturer's responsibility may be held to go beyond mere literal compliance with the statutory law and meet more stringent requirements.

I also pointed out that courts have found manufacturers in violation of the requirement of adequacy for such reasons as failure to communicate the level of the danger, failure to display the warning in a reasonable location, failure to give sufficient information about how to avoid the danger, and failure to preserve the integrity of the warning by including statements that nullify its impact. The findings from the research reported here suggest that if cigarette manufacturers were held to more rigid requirements with regard to adequacy and warning, there would be reason to hold them liable for such failure.

However the law develops with respect to this issue, there seem to be compelling reasons to understand better how warnings are perceived and processed by individuals. If it develops that cigarette manufacturers are liable for adequacy of warning, then the tobacco industry will want to protect its interests by developing warning labels that are at least legally adequate. If, on the other hand, Congress continues to mandate the wording and design of cigarette package warnings, information about how to design and word

warning labels should be available to it. And as noted earlier, the whole field of product liability maintains a strong interest in the general warning issue (see also Johnson, Chapter 10 this volume).

5.2. Classification of Warnings

Although it seems true, as I noted earlier, that the formalization of semantic and syntactic rules for warnings is difficult, it now seems possible that linguistic and human factors research will eventually permit us to discover some generalizations about the nature, structure, and function of warnings.

With respect to the possible distinction between Searle's categorical (or informational) warnings and hypothetical (or predictive) warnings, the evidence from these experiments is inconclusive. Although it may be true that the distinction is in some way psychologically real, further research is needed to establish that. All I have established is that my respondents seemed to recognize hypothetical warnings with great consistency.

In particular, it is unclear whether respondents' consistency of response implicates any syntactic factors or whether it is purely a matter of logical structure. Let us consider the four hypothetical warnings once again.

1. Give me your money, or I'll shoot!
2. If you choose not to remain silent, anything you say or write can and will be used against you in a court of law.
3. When you go and snitch to anyone that we had anything to do with this, you'll find a snitch tattoo on your forehead.
4. Take heed, sweet soul.

First of all, it is clear that hypothetical warnings need not lexically express both logical operators IF and THEN, or even one of them. In 2 and 3, which have the most canonical form for hypothetical warnings of the four, the "then" is not expressed overtly, whereas "when" substitutes for "if" in 3. In 4, still more is unexpressed—the entire second (consequent) clause—which the addressee/hearer must then infer from the context. And in 2, the typical "If X, then Y" structure is replaced by a functionally equivalent structure, consisting of an imperative conjoined to a declarative—without any overt markings of the "if/then" interpretation that the structure nevertheless evokes.

Although 1 is interpretable as a warning, it is important to note that in this case, at least, it is not simply the syntactic structure that signals its illocutionary force, because this structure may be used for other kinds of speech acts that, like warnings, make a statement about the future intent of the speaker (e.g., threats—"Halt or I'll shoot," or promises—"Get an *A* in the course, and I'll give you a hundred dollars"). On the other hand, even an explicit "if/then" structure is not sufficient to trigger a warning interpretation, because that structure may be used in speech acts quite different from warnings (e.g.,

statements of natural laws—"If you heat water to 212 degrees, then it will boil").

Because the data suggest that there is no specific syntactic structure that is either necessary or sufficient to signal a hypothetical warning in these experiments, further research is necessary to decide on what basis subjects have categorized these warnings as hypothetical. Such research must then be used in further explorations of the validity of Searle's distinction between categorical and hypothetical warnings.

5.3. Summary of Findings; Conclusions

With respect to my general hypotheses, the first four were supported by the findings in Experiments 2 and 3. Responses and comments suggest that the modal auxiliaries "may" and "can" tend to weaken warnings; that disease-specific warnings are perceived as stronger than non-disease-specific warnings; that unusual syntax, as in the *double -ing* construction, may render warnings ineffective; and that incomplete warnings (*Cigarette Smoke Contains Carbon Monoxide*) are perceived as silly. There is some evidence that the inclusion of limited-target warnings in a rotational series may have the effect of weakening the overall effect of the series. There is also some evidence to suggest that people process general warnings very individually, in terms of such factors as their own values and risk-taking behavior. Much additional research is needed, however.

As stated at the outset of this chapter, my specific goal in the original trial context was to offer some linguistic and human factors evidence about the adequacy of the 1970 cigarette package warning with respect to ill-health consequences, specifically the probably negative cardiovascular effects of cigarette smoking. In order to be able to do that, I addressed eight specific questions that I had formulated on the basis of my preliminary research. The answers to those questions follow.

Q1. Can warnings be identified by the presence of specific semantic, lexical, syntactic, or other characteristics? If so, what are they?

A1. Though I did not conduct research into how individuals define the category of warnings or how they recognize that statements constitute warnings, my respondents did seem to accept as warnings statements that make reference to a future event or state that is, at least in the opinion of the speaker or writer, not in the best interests of the addressee. And they occasionally objected to labeling as *warnings* statements that clearly lack those characteristics. Although some warnings can be identified by the presence of some specific logical and lexical characteristics, many others cannot. Conventional labels (e.g., *Warning*) and words like *POISON* are

among the lexical items that sometimes characterize warnings. Further, respondents' comments suggested that they expect printed warnings on consumer products to be conspicuous with respect to such factors as type size, color contrast, and position. In other words, respondents believe that statements that otherwise constitute warnings should not be nullified or rendered ineffective by factors rendering them difficult to read.

Q2. Can warnings be meaningfully classified? If so, on what bases? Do consumers respond to or evaluate warnings on such bases?

A2. Warnings can be classified in at least one meaningful way. As noted, the evidence was that the distinction between categorical and informational warnings and hypothetical or predictive warnings appears to be in some way empirically discoverable. Respondents seemed to recognize hypothetical warnings with great consistency, apparently responding to explicit or implicit IF–THEN logical structure very consistently.

Q3. Do warnings differ by degree? Do consumers classify warnings as strong versus weak?

A3. There is some evidence that, with the exception of conventional categorical words like *POISON*, English-speaking adults rank hypothetical warnings as stronger than categorical warnings. Further, external factors (type size, type color, position, etc.) play a role in the way respondents rank warnings. Warning labels that contain qualifiers (especially the modal auxiliaries "may" and "can") are ranked as weaker than other warnings.

Q4. Do the statements at issue here constitute warnings? If so, precisely what do the statements appear to warn against?

A4. The labeled warnings on cigarette packages manufactured in the United States are accepted as warnings by most respondents, though, upon being questioned, some respondents pointed out that the two new rotating warnings mentioning carbon monoxide and the advantage to *quitting smoking now* do not really warn. The different warnings are seen as warning different addressees against different dangers.

Q5. Do warnings differ by degree, that is, do consumers classify warnings as strong versus weak?

A5. Warnings differ by degree, that is, consumers classify warnings as strong or weak, effective or ineffective, inclusive or exclusive. Strong warnings are characterized by the following features, no one of which must be present, but all of which are frequently present in strong warnings:

1. They are often formulated either as hypothetical warnings or as powerful fear provokers, like *POISON*.

2. They are also frequently marked by an absence of qualifiers (e.g., *may, could*) and by the fact that they mention specific possible negative consequences of behavior.
3. They are generally conspicuous, for example, they are not nullified by being rendered practically invisible or easy to ignore because of type size, type color, or type position.
4. They are written in simple syntax and in ordinary, everyday language.
5. They often contain specific information about precise consequences.

Weak warnings lack one or more of these characteristics.

Q6. Are the federally mandated cigarette package warnings strong or weak? If they are weak, how could they be made strong?

A6. The federally mandated cigarette package warnings vary considerably in strength, ranging from fairly strong to silly. They could have been made stronger by following suggestions made by the FTC or by following suggestions made in market research surveys and known to the Tobacco Institute in the 1970s (Myers, Iscoe, Jenning *et al.*, 1980).

Q7. Are the warnings at issue adequate to warn consumers of possible negative effects of cigarette smoking, particularly with regard to the potentially negative cardiovascular effects of cigarette smoking? If they are not, could they have been?

A7. Some of the warnings at issue appear adequate to warn consumers of the general possible ill effects of cigarette smoking, with the exception that none of them warns of possible addiction. My research revealed no likelihood that the warning labels serve to warn against the possibly negative cardiovascular effects of cigarette smoking. Looking for such a connection was an original goal of the research.

These findings thus constitute evidence for the existence of a variety of objective criteria by means of which the relative adequacy of warnings on cigarette packages can be assessed. However, my research also shows that these objective criteria are for the most part not characteristic of present or past warnings on cigarette packages. The warnings I tested demonstrate clear inadequacies with respect to both linguistic and extralinguistic factors characteristic of strong warnings.

In particular, federally mandated cigarette package warnings display characteristics of weak warnings: (1) qualifying language (e.g., the modal auxiliaries *may* and *can*), (2) unusual syntax (e.g., the *double -ing* construction as in *Quitting smoking now*), and (3) technical and semitechnical vocabulary (e.g., *fetal injury, carbon monoxide*). The warnings lack significant information

(What are the precise dangers? Who will be affected? To what extent?) The warning labels are hard to read because of their position on the side of the package, their small type size, and the fact that they often appear in hard-to-read color combinations (e.g., gold on red).

Also, space limitations constitute a very serious problem in choosing informative wording. As mentioned, one of my graduate students reported that some pregnant women thought that *Low Birth Weight* was a desirable result of smoking and that it was unclear whether that was because they interpreted the statement to mean that *their* weight would be lower at birth or that having a baby weighing less would be desirable. Attempting to propose alternative wordings makes it clear that some problems inherent in the warning would take more words to clear up than there is room for on the package. An informationally adequate warning might read thus: "Smoking by pregnant women may cause injury to the baby before birth, as well as dangerous health problems resulting from the baby's being born prematurely." Here and elsewhere in the rotational system, brevity is accomplished by the use of typically opaque nominalizations (e.g., *fetal injury, birth weight*) made even more opaque by the omission of crucial information for which there is simply no room. This situation-required brevity is a serious obstacle to the adequacy of these warnings.

The rotating warnings present the additional problem that differences in the strength of individual warnings and in the breadth of target populations appear to have the effect of weakening stronger or more inclusive warnings. There is some evidence that two of the present rotating warnings have the effect of weakening the effectiveness of the one disease-specific warning in current use.

6. RECOMMENDATIONS

On the basis of the research reported here, I recommend that the following steps be taken in designing warning labels for consumer products, particularly where health hazards are to be warned against and where consumers will be expected to process the information given them in different ways given differing values and risk-taking stances:

1. Either formulate the warnings as hypothetical warnings or use strong conventional warning labels like *POISON*.
2. Avoid unnecessary qualifying language, for example, the modal auxiliaries *may* and *can*.
3. List specific undesirable consequences of unsafe behavior.
4. Make the warnings conspicuous in all ways, for example, color contrast, type size, and position on product.

5. Write the warnings in simple syntax and in ordinary vocabulary.
6. Include specific information about negative consequences on each label in a rotational series.
7. Do not narrow the target population by addressing specific labels to different portions of that population (e.g., pregnant women).
8. When considering the use of rotating warnings, consider that differences in the strength of individual warnings may have the effect of weakening stronger warnings.
9. Field-test all proposed warnings. (This step would appear to go without saying, but, given the history of proposed federally mandated warnings, it is clear that it does not.)

7. DIRECTIONS FOR FUTURE RESEARCH

The questions addressed here, all revolving around notions of adequacy of warning, exist in a political/economic arena in which it is often tempting to dismiss the need for and advantages of research in favor of simple or simplistic answers to the questions. During the 3 years that I have been engaged in the research described here, I have participated in many discussions, some heated, with individuals who dismiss the need for the research described here on the basis that individuals are responsible for their own fate and that any discussion of adequacy of warning is beside the real point, which is that manufacturers should not bear the financial burden for individuals who choose to engage in risky behavior.

Others have taken the almost diametrically opposed point of view and have argued that the courts must do what Congress is too cowardly to do, that is, assess the monetary damages to society of cigarette smoking to tobacco product manufacturers in the form of large awards to plaintiffs and their families.

Both these arguments have important implications for the legal questions dealt with in this chapter, particularly the very important one of whether Congressional promulgation of cigarette package warnings has preempted common-law requirements of adequacy of warning. Neither argument, however, addresses at all the indisputable facts that (1) warnings, like other speech acts, vary in communicative effectiveness and (2) it is possible to learn a great deal about the nature of the factors that promote communicative effectiveness. Although it may be questionable whether printed warnings labels are the most effective way of conveying certain kinds of health information to consumers, it seems to be true that some warning labels are superior to others. So long as we are going to require cigarette manufacturers to print warning labels on cigarette packages for the purpose of informing consumers of the

probable negative consequences of smoking cigarettes, then we should also require that those warnings communicate as effectively as possible.

In the interest of continuing to advance our state of knowledge with respect to warning effectiveness, I plan to complete at least three additional research phases. In the first, I plan to explore the line of research of Kreckel's 1981 study suggesting that "[t]here is no 'natural' common core for 'warning' " (p. 87) and that ' "warnings in general' do not exist" (p. 87) (see footnote 3). I will do so by asking respondents to identify speech acts in tape-recorded conversations (using the same kind of brief instruction I described in connection with the written questionnaire). This research will have the primary goal of discovering to what extent subjects can identify warnings consistently.

In the second proposed research plan, I shall collect much larger samples of paired comparisons using many more warnings than were used in the research described here, so that my findings will be amenable to scaling analysis.[13] This research will have the primary goal of discovering the extent to which respondents display consistency in selecting the stronger of pairs of warnings. It will also shed light on the extent to which certain factors are characteristic of warnings rated as strong.

Finally, in a third phase, I plan to collect information about how respondents process cigarette warnings in the light of their own individual risk assessment techniques by making use of a computerized system, ARK, for studying the ways in which people assess risk. ARK was designed by researchers at the Oak Ridge National Laboratory for the purpose of collecting detailed information about how users assess risk in many specific domains, including cigarette smoking. Its use requires considerable computational and statistical sophistication because of the nature of the questions asked by the system. However, it is the most comprehensive system I am familiar with for gathering the kind of information it gathers. My plan is to instruct colleagues and students in its use and collect information over several years. This research will have the primary goal of discovering something about the relationship between the risk-assessing techniques of individuals and the ways in which they process tobacco product warnings.

[13] Among those I will add are two required on cigarette packages by the governments of Canada and the United Kingdom. Both countries mandate the use of cigarette package warnings that would, on the basis of my research findings, be rated as stronger than some of those in use in this country. The Canadian warning reads, "WARNING: Health and Welfare Canada advises that danger to health increases with amount smoked—avoid inhaling." Arguably, that gives permission to smoke; however, it also identifies the precise source of greatest health danger. The one mandated in the United Kingdom reads "DANGER: Government Health WARNING: Cigarettes Can Seriously Damage Your Health." In spite of the presence in the warning of the modal auxiliary "can," I think the warning would, on the basis of its simplicity and label, test strongly. We still have much to learn about various combinations of factors, syntactic and otherwise.

The sequence of research steps outlined should expand our knowledge of three questions, the first two of which were already discussed: (1) What does it mean to be warned? (2) What factors enhance the effectiveness of a warning? (3) How do individuals' own risk assessment techniques interact with warnings to promote or constrain behavioral response to warnings? Such knowledge may have the effect of enhancing our ability to adequately inform consumers of health risks and thus better enable all of us to make informed choices about risky behavior.[14]

8. REFERENCES

A glossary of tobacco terms. (1985). *New York State Journal of Medicine, 85,* 297–298.

Abramson, J. (1986, November 10). "Battle lines drawn in cigarette ad fight." *Legal Times 9,* p. 1.

Black's Law Dictionary (5th ed. 1979).

Breznitz, S. (1984). *Cry Wolf: The psychology of false alarms.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Blasi, V., & Monaghan, H. P. (1986). The first amendment and cigarette advertising. *Journal of the American Medical Association, 256,* 502–509.

Cullingford, R., Da Cruz, L., Webb, S., Shean, & Jamrozik, K. (1988). Legibility of health warnings on billboards that advertise cigarettes. *The Medical Journal of Australia, 148,* 336–338.

Davis, R. M., & Kendrick, J. S. (1989). The Surgeon General's warnings in outdoor cigarette advertising: Are they readable? *Journal of the American Medical Association, 261,* 90–94.

De Greene, K. B. (1970). *Systems psychology.* New York: McGraw-Hill.

Deposition of Dr. Roger L. McCarthy. (1987, July 2). Frank Cusimano, Jr., Plaintiffs, vs. American Honda Motor Company, *et al.*, Defendants, in the Superior Court of the State of California in and for the County of San Diego, Docket No. 499782, Vol II.

Fischer, P. M., Richards, J. W., Berman, E. J., & Krugman, D. M. (1989). Recall and eye tracking study of adolescents viewing tobacco advertisements. *Journal of the American Medical Association, 261,* 84–89.

Fraser, B. (1975). Warning and threatening. *Centrum, 3,* 169–180.

Gidmark, D. (1986, December 1). A tobacco case activist predicts success by end of '87. *The National Law Journal, 1,* 8–9.

Kreckel, M. (1981a). Where do constitutive rules for speech acts come from? *Language and Communication, 1,* 73–88.

Kreckel, M. (1981b). *Communicative acts and shared knowledge in natural discourse.* London: Academic Press.

Labov, W. (1966). *The social stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). The social stratification of (r) in New York City department stores. In W. Labov, *Sociolinguistic patterns* (pp. 43–69). Philadelphia: University of Pennsylvania Press.

Lee, J. D. (1986, Winter). Three stages of tobacco litigation. *Trial Diplomacy Journal, 8,* 18–23.

---

[14] Earlier and shorter versions of this paper were delivered at The University of Tennessee (February 20, 1986), Duke University (March 25, 1986), and at the 36th meeting of the Southeastern Conference on Linguistics at Georgetown University (March 27, 1987). Requests for reprints should be directed to the author at the English Department, The University of Tennessee, Knoxville, Tennessee 37996-0430 (BITNET%"DUMASB@UTKVX").

Lehto, M. R., & Miller, J. M. (1986). *Warnings, Volume I: Fundamentals, design, and evaluation methodologies*. Ann Arbor: Fuller Technical Publications.

McCormick, E. J. (1970). *Human factors in engineering and design* (3rd ed.). New York: McGraw-Hill.

Miller, J. M., & Lehto, M. R. (1987). *Warnings: Volume II: Annotated bibliography with topical index*. Ann Arbor: Fuller Technical Publications.

Murrell, K. F. H. (1969). *Man in his working environment*. London: Chapman and Hall.

Myers, M. L., Iscoe, C., Jennings, C., *et al.* (1981). *Federal Trade Commission staff report on the cigarette advertising investigation*. Washington, DC: Federal Trade Commission.

Office of Smoking and Health, U.S. Department of Health, Education and Welfare (1964). *Smoking and health: A report of the Surgeon General*. Washington, DC: Author.

Restatement (Second) of Torts §§ 338, 402a (1965).

Robinson, G. H. (1977). Human performance in accident causation: Toward theories on warning systems and hazard appreciation. Washington, DC, *Third International System Safety Conference*, Oct. 17–21, pp. 55–69.

Samfield, M. (1980). *Research and manufacturing in the U.S. cigarette industry*. New York: Lockwood.

Schwartz, V. E., & Driver, R. W. (1983). Warnings in the workplace: The need for a synthesis of law and communication theory. *Cincinnati Law Review, 52*, 38–83.

Searle, J. (1969). *Speech Acts: An essay in the philosophy of language*. London: Cambridge University Press.

Weinstein, A. S., Twerski, A. D., Piehler, H. R., & Donaher, W. A. (1978). Chapter 5: Warnings and disclaimers. *Products liability and the reasonably safe product* (pp. 60–74). New York: John Wiley.

9. CASES CITED

American Optical v. Weiderhamer, 404 N.E.2d 606 (Ind.App 4 Dist. 1980).

Bituminous Casualty Corp. v. Black and Decker Corp. 518 S.W.2d 868 (Tex.Civ.App. 5 Dist., 1974).

Ford Motor Co. v. Nowak, 638 S.W.2d 582 (Tex.App. 13 Dist., 1982).

Griggs v. Firestone Tire and Rubber Co., 513 F.2d 851 (8th Cir. 1975).

Mullane v. Central Hanover Bank & Trust Co., 339 U.S. 306 (1950).

Nissen v. Miller, 44 N.M. 487, 490, 105 P.2d 324, 326 (1940).

Roysdon v. R. J. Reynolds Tobacco Co., 623 F. Supp. 1189 (E. D. Tn. 1986).

Roysdon v. R. J. Reynolds Tobacco Co., 849 F.2d 230 (6th Cir. 1988), *rehearing denied* 1988.

Wallinger v. Martin Stamping and Stove Co., 93 Ill.App.2d 437, 236 N.E.2d 755 (Ill.App. 1968).

10. STATUTES AND REGULATIONS CITED

Comprehensive Smokeless Tobacco Health Education Act of 1986, 15 U.S.C. (Supp. 1988).

Comprehensive Smoking Education Act of 1984, 15 U.S.C. (1984).

Federal Cigarette Labeling and Advertising Act of 1965, 15 U.S.C. (1965).

Public Health Cigarette Smoking Act of 1969, 15 U.S.C. (1970).

The Trade Regulation Rule for the Prevention of Unfair or Deceptive Advertising and Labeling of Cigarettes in Relation to the Health Hazards of Smoking, 29 Fed. Reg. §§ 8324, 8325, 8326 (July 2, 1964).

COMMENTS

FINANCIAL INSTITUTION DEPOSIT INSURANCE—DIRECTORS' AND OFFICERS' LIABILITY INSURANCE POLICIES—PUBLIC POLICY REGARDING REGULATORY EXCLUSIONS

I. INTRODUCTION

In response to the Depression of the 1930s, which was exacerbated by the collapse of banks following runs by depositors, quasi-federal agencies were developed to stabilize the banking system. The Federal Deposit Insurance Corporation (FDIC), founded in 1933, was to insure deposits for national banks and members of the Federal Reserve[1] up to a statutorily determined limit. The government also established the Federal Savings and Loan Insurance Corporation in 1934 to insure deposits in savings and loans.[2] These deposit insurance funds did in fact restore faith in the soundness of the banking system.[3]

In the mid-1980s, however, there was a significant increase in the number of financial institution failures each year.[4] The magnitude of the number and size of the institutions closed because of insolvency shook the entire deposit insurance system.[5] The Federal Savings and Loan Insurance Corporation (FSLIC) became insolvent in 1986 and was later abolished and its duties merged into the FDIC.[6] Congress

---

1. FDIC, SYMBOL OF CONFIDENCE 3 (1990); EDWARD L. SYMONS, JR. & JAMES J. WHITE, BANKING LAW 570 (2d ed. 1984) [hereinafter SYMONS & WHITE].

2. John R. Cranford, *Deposit Insurance: A History*, CONG. Q. (Feb. 17, 1990).

3. FDIC, *supra* note 1, at 9; SYMONS & WHITE, *supra* note 1, at 570.

4. 1990 FDIC ANN. REP. 77.

5. 1989 FDIC ANN. REP. vii-xi.

6. Cranford, *supra* note 2. Enactment of the Financial Institutions Reform, Recovery, and Enforcement Act (FIRREA) in 1989 eliminated the FSLIC and established the Office of Thrift Supervision to supervise healthy savings associations and the Resolution Trust Corporation (RTC) to liquidate failed savings and loans. The RTC and the Savings Association Insurance Fund were added to the FDIC's prior functions of managing the Deposit Insurance Fund (renamed by FIRREA the Bank Insurance Fund) and liquidating failed banks. FDIC, SYMBOL OF CONFIDENCE 9-11; 1989 FDIC ANN. REP. 85.

has passed legislation in recent years to keep the FDIC afloat.[7]

Many banks that failed were the holders of directors' and officers' liability insurance policies (D&O policies) that covered bank losses because of dishonest acts of employees. FDIC or FSLIC recovery under these policies recouped at least some of the losses of the bank and thereby lessened the expenditure necessary from deposit insurance funds.

Insurers, however, began to add limiting language to the D&O policies to reduce their liability exposure. One of the limitations developed was a "regulatory exclusion"[8] that either terminated coverage upon the appointment of a regulatory agency as receiver of a financial institution[9] or denied coverage for claims brought "by or on behalf of" a regulatory agency.[10] Under this type of policy, only claims filed with the insurer before the bank failure were recoverable.

Regulatory agencies repeatedly have challenged these exclusions in courts, with mixed results. Initially, some courts upheld the validity of these exclusions,[11] while other courts found them void as against public policy.[12] Although Congress acknowledged an awareness of the issue in 1989 while amending 12 U.S.C. § 1821,[13] they decided, for unknown reasons, to allow the judicial system to continue grappling with the controversy.[14]

In 1990, the Court of Appeals for the Sixth Circuit in *FDIC v. Aetna Casualty and Surety Co.* found there was no clear manifestation of public policy against regulatory exclusions and held them

---

7. 12 U.S.C. § 1824(a) was amended on August 9, 1989 to increase the borrowing authority of the FDIC from $3,000,000,000 to $5,000,000,000. 12 U.S.C. § 1824 (Law. Co-op. Supp. 1991). This limit was raised further in December of 1991 to $30,000,000,000. Federal Deposit Insurance Corporation Improvement Act of 1991, Pub. L. No. 102-142, § 101, 105 Stat. 2236 (1991).

8. Johnston & Glancz, *Current Issues in Officer/Director Liability, Indemnification, and Insurance Coverage*, 51 WASH. FIN. REP. (BNA) 811 (Nov. 7, 1988); Peter G. Weinstuck, *Directors and Officers of Failing Banks: Pitfalls and Precautions*, 106 BANKING L.J. 434, 449 (Sept.-Oct. 1989).

9. "This bond shall be deemed terminated or canceled as an entirety . . . immediately upon the taking over of the Insured by a receiver or other liquidator or by State or Federal officials . . . ." FDIC v. Aetna Casualty & Sur. Co., 903 F.2d 1073, 1075 (6th Cir. 1990).

10. "[T]he company shall not be liable . . . [for] any claim . . . brought by or on behalf of . . . [any] regulatory agency . . . ." FSLIC v. Oldenburg, 671 F. Supp. 720, 722 (D. Utah 1987).

11. Sharp v. FSLIC, 858 F.2d 1042 (5th Cir. 1988); FSLIC v. Transamerica Ins., 705 F. Supp. 1328 (N.D. Ill. 1989); Continental Casualty Co. v. Allen, 710 F. Supp. 1088 (N.D. Tex. 1989).

12. Branning v. CNA Ins. Co., 721 F. Supp. 1180 (W.D. Wash. 1989); FSLIC v. Aetna Casualty & Sur. Co., 701 F. Supp. 1357 (E.D. Tenn. 1988); FSLIC v. Oldenburg, 671 F. Supp. 720 (D. Utah 1987).

13. H.R. REP. No. 54, 101st Cong., 1st Sess., pt. 1, *reprinted in* 1989 U.S.C.C.A.N. 86,127.

14. *See infra* p. 320.

valid and enforceable.[15] Later cases have upheld regulatory exclusions in the face of public policy challenges primarily by following that decision.[16] As shall be discussed later, however, while the court in *FDIC v. Aetna* could reach the decision it did, its analysis was incomplete and misleading. Therefore, the issue has yet to be satisfactorily resolved.

II. PUBLIC POLICY

The Supreme Court faced the issue whether a specific contract or type of contract was void as against public policy as early as 1900.[17] While recognizing some contracts were void for this reason, the Court, in *Baltimore & Ohio Southwestern Railway v. Voigt*, cautioned lower courts that one important portion of the liberty enjoyed by American citizens was the right to enter into private contracts.[18] The Court said the "usual and most important function of courts of justice" was to enforce contracts rather than to allow a party to escape its contractual obligations on the pretext of public policy.[19] The rules enabling a court to void a contract as against public policy are not to be arbitrarily extended because public policy

---

15. FDIC v. Aetna Casualty & Sur. Co., 903 F.2d 1073 (6th Cir. 1990).

16. St. Paul Fire & Marine Ins. Co. v. FDIC, 765 F. Supp. 538, 549-50 (D. Minn. 1991); American Casualty Co. v. Baker, 758 F. Supp. 1340, 1345-47 (C.D. Cal. 1991); Gary v. American Casualty Co., 753 F. Supp. 1547, 1553 (W.D. Okla. 1990) (a somewhat more thorough analysis of statutes and legislative history than *FDIC v. Aetna*).

17. Baltimore & O.S.W. Ry. v. Voigt, 176 U.S. 498 (1900). In previous cases the Court had ruled railroad companies acting as a common carrier for passengers for hire could not stipulate that in purchasing a ticket a passenger was contracting to relieve the common carrier from liability for losses or injuries caused by their negligence. *Id.*

18. *Id.* at 505. The Court premised voiding waivers of liability in favor of common carriers on two public policy principles. *Id.* at 506. First, the importance the law attaches to human life and personal safety and the reduction in safety that would occur if the waivers were enforceable. *Id.* The diligence and care exercised by common carriers would decline should they be protected from liability for negligence. *Id.* Second, the carrier and customer do not stand on a footing of equality because the customer is only one of a million customers and has no real bargaining power. *Id.* The carrier cannot dictate terms of liability and have them enforced. *Id.*

19. *Id.* at 505. In *Baltimore & Ohio*, the Court declined to void the contract in question on the basis of an asserted public policy. 176 U.S. at 520. The plaintiff was an employee of an express company and, in gaining employment, had contracted to waive any claim of liability against the employer. *Id.* at 500. He also had ratified an agreement between the employer and the railroad company that relieved the railroad of liability to any employees of the express company. *Id.* at 501. The Court held plaintiff was not constrained to enter the contracts in question but entered into them freely and voluntarily, obtaining the benefit of securing employment, and that they did not contravene public policy. *Id.*

primarily requires that the freedom of contract be protected. Contracts freely and voluntarily entered by competent adults should be enforced by the courts. The paramount public policy is for courts not to interfere lightly with the freedom of contract.[20]

Courts must strike a delicate balance when the merits of voiding a contract as against public policy are weighed against the sanctity of the freedom to enter enforceable contracts.[21] In *Steele v. Drummond*[22] the Supreme Court recognized,

> the meaning of the phrase "public policy" is vague and variable; there are no fixed rules by which to determine what it is. It has never been defined by the courts, but has been left loose and free of definition. . . . It is only in clear cases that contracts will be held void. The principle must be cautiously applied to guard against confusion and injustice. Detriment to the public interest will not be presumed where nothing sinister or improper is done or contemplated. [citations omitted].[23]

In 1931, the Supreme Court in *Twin City Pipeline Co. v. Harding Glass Co.* again considered a case in which one of the parties tried to have its contract declared invalid because of an alleged violation of public policy.[24] In determining the validity of the claim, the Court reaffirmed its previously expressed views: (1) contract enforceability is not to be denied arbitrarily; (2) there are no set rules to determine which contracts are repugnant to public policy; (3) contracts in contravention of public policy should be held not enforceable only in cases plainly within the reasons upon which that doctrine rests; and (4) only a *dominant* public interest should allow one party to escape from an otherwise valid contract.[25] In determining whether public policy renders a contract void, the courts are also to consider the "Constitution, laws, and judicial decisions of [the jurisdiction] as well [as] the applicable principles of the common law."[26] A contract may not be invalidated without "a clear showing that some

---

20. *Baltimore & Ohio*, 176 U.S. at 505.

21. *See* Steele v. Drummond, 275 U.S. 199, 205 (1927).

22. *Id.* at 199.

23. *Id.* at 205.

24. Twin City Pipe Line Co. v. Harding Glass Co., 283 U.S. 353 (1931). The defendant, Harding Glass Company, in settlement of litigation, had entered a contract whereby it was required to purchase all of its gas from the Twin City Pipe Line Company as long as the pipe line company could adequately supply its needs. In return, the pipe line company agreed to build an additional line to supply gas to the glass company. The agreement was adhered to by both parties for over two years, at which time the defendant began receiving gas from a newly completed gas line built by a subsidiary company. *Id.* at 356. In defense of its actions, the defendant alleged the contract was void because of a supposed violation of the provision in the Arkansas Constitution which prohibited monopolies. *Id.* at 357.

25. *Id.* at 356-57 (emphasis added).

26. *Id.* at 357.

definite public detriment will probably result from its performance."[27]

Courts are to determine public policy by reference to the laws and legal precedents and not from general considerations of supposed public interests.[28] In making their determination, the statutory provisions are of considerable importance.[29] In addition, determining whether a contract is enforceable requires "consideration of the broad purposes of relevant statutes and the probable effect upon this."[30] In at least two public policy cases, the end Congress intended to accomplish was treated as the controlling factor.[31]

Litigants, however, frequently raise "public policy" arguments in situations where no statutes exist specifically addressing the matter before a court.[32] Other cases arise when Congress has prohibited a specific type of contract but the statute does not encompass the contract before the court.[33] Here also, the purpose of Congress, the "vice" to be avoided, is critical in the court's consideration.[34]

In *Muschany v. United States*,[35] decided in 1945, the problem Congress had legislated against was not present in the contract before the Court.[36] Although there was a statute prohibiting cost-plus-a-percentage-of-cost contracts, the contract before the Court was contingent and thus not barred by the statute. While the Court acknowl-

---

27. *Id.* at 358.

28. United Paperworkers Int'l Union v. Misco, Inc., 484 U.S. 29 (1987); Muschany v. United States, 324 U.S. 49 (1945); W.R. Grace & Co. v. Local Union 759, Int'l Union of the United Rubber Workers, 461 U.S. 766 (1983); Vidal v. Mayor of Philadelphia, 43 U.S. (2 How.) 127 (1844).

29. *Twin City*, 283 U.S. at 357.

30. A.C. Frost & Co. v. Coeur D'Alene Mines Corp., 312 U.S. 38, 44-45 (1941).

31. *Id.* at 45.

32. *See, e.g.*, Baltimore & O.S.W. Ry. v. Voigt, 176 U.S. 498 (1900).

33. Muschany v. United States, 324 U.S. 49 (1945). In *Muschany*, the Court considered whether a government contract violated a provision of a federal statute prohibiting the War Department from using cost-plus-a-percentage-of-cost contracts. *Id.* at 51. During a war emergency, the Department contracted with a civilian who was to secure option contracts for the purchase of land for an ordnance plant. *Id.* at 51-52. The contract provided that the seller would pay the civilian a five percent commission. *Id.* Viewing this arrangement as distinctly different from the prohibited cost-percentage contracts, the Court upheld the contract. *Id.* at 62, 69. The Court found the "vice of cost-plus contracts is not inherent in the vendor's contract" because the Government was already bound to pay future costs and the incentive for the contractor to inflate costs to increase his profits was not present. *Id.* at 62-63. The final cost was already fixed and known to the Government when it accepted the option. *Id.* at 63. Congress "did not care how the contractor computed his fee or profit so long as [it] was finally and conclusively fixed in amount at the time when the Government became bound to pay it by its acceptance of the bid." *Id.* at 62.

34. *See id.* at 61-69.

35. 324 U.S. 49 (1945).

36. *See supra* note 33 and accompanying text.

edged Congress could, if desired, ban contingent fee governmental contracts also, it stated "until it does we cannot say that they are contrary to public policy, any more than we could say cost-plus contracts are contrary to public policy in the absence of legislation to that effect."[37] Thus, although cost-plus contracts were abused to increase the profit of a vendor at the expense of the Government,[38] and ultimately the taxpayer, the Court would invalidate them only after Congress passed legislation prohibiting cost-plus contracts. Although similar abuses were possible under the contingent contracts at issue in *Muschany*,[39] the Court would only void them on public policy grounds if Congress specifically prohibited contingent contracts.[40] It is important to the public, said the Court, that good faith contracts made between the United States and citizens

> shall not be lightly invalidated. Only dominant public policy would justify such action. Without a plain indication of that policy through long governmental practice or statutory enactments, or of violations of obvious ethical or moral standards, [the] Court should not assume to declare contracts of the [Government] contrary to public policy. The courts must be content to await legislative action.[41] It is Congressional enactments that determine public policy.[42]

Muschany reveals that courts must look to more than the words of the statute in determining whether Congress has expressed a public policy.[43] When the Fourth Circuit was faced with a request to hold a contract void, it acknowledged that courts will not generally enforce contracts "made in derogation of statutes designed to protect the public."[44] That court, however, also recognized, "[b]efore holding contracts unenforceable on public policy grounds, such as illegality, . . . it is necessary to determine just what acts are forbidden. . . . Courts are averse to holding contracts unenforceable on the ground of public policy unless their illegality is clear and certain."[45]

---

37. 324 U.S. at 65.
38. *Id.* at 61-62.
39. *Muschany*, 324 U.S. at 65-66. The Court recognized the temptation existed for the soliciting agent to encourage the seller to raise his price, thus resulting in a higher commission for the agent. *Id.* at 72 (5-3 decision) (Black, J., dissenting) (arguing that only by acting to the detriment of the Government could the agent act for his own financial advantage and "[i]t was to protect the public from the dangerous tendency of the excesses of just such contracts that Congress" prohibited cost-plus contracts).
40. *Id.* at 65.
41. *Id.* at 66-67.
42. *Muschany*, 324 U.S. at 68.
43. *Id.* at 61-62.
44. Smithy Braeden Co. v. Hadid, 825 F.2d 787, 790 (1987).
45. *Id.*

III. REGULATORY EXCLUSIONS

A. Initial Cases Determining Enforceability in D&O Policies

Within this general analytical framework governing the enforceability of contracts in the face of a public policy challenge, courts have considered the specific issue whether regulatory exclusion clauses contained in D&O insurance policies are enforceable against government agencies succeeding to the rights of failed financial institutions. The results have been mixed.

1. Unenforceable—Violation of Public Policy

In *FSLIC v. Oldenburg*,[46] the FSLIC sought a declaratory judgment to establish the existence of coverage under the D&O indemnity insurance.[47] The exclusionary provision, which stated the insurer was not liable for loss in connection with a claim brought by or for any regulatory agency,[48] was said to bar recovery by the FSLIC. The FSLIC challenged the provision's enforceability based on a public policy argument.[49]

The court in *Oldenburg*, the United States District Court for Utah, found a strong policy expressed by federal statutes and regulations. This public policy gave the FSLIC, as receiver, all the rights and claims the original insured would have had.[50] The court rejected

---

46. 671 F. Supp. 720 (D. Utah 1987).

47. *Id.* Federal Insurance Company (Federal) issued a D&O policy to State Savings. *Id.* at 722. The FSLIC acquired the policy when it was named receiver of State Savings. *Id.* at 722 n.1. Evidently the FSLIC wanted to sue the former directors and officers of State Savings. However, the FSLIC wanted to be certain Federal would be forced to abide by the policy and pay any judgment against the directors and officers. Presumably, if Federal could successfully deny liability, the FSLIC would have to pursue the directors and officers based on their personal net worth.

48. Endorsement No. 2 provided:

> It is understood and agreed that the company shall not be liable to make any payment for loss in connection with any claim made against the directors or officers based upon or attributable to any claim, action or proceeding brought by or on behalf of the Federal Home Loan Bank Board, any other similar organization, or any other national or state bank, regulatory agency, whether such claim, action or proceeding is brought in the name of such regulatory agency or by or on behalf of such regulatory agency in the name of any other entity.

Oldenburg, 671 F. Supp. at 722.

49. *Id.* The Court found it unnecessary to consider FSLIC's separate ambiguity argument. *Id.*

50. *Id.* at 723. Pursuant to the statutory authority granted in 12 U.S.C. § 1464(d)(11) (Supp. 1987), the Federal Home Loan Bank Board (FHLBB) adopted regulations pertaining to the powers and rights of receivers. These regulations stated:

> [A] receiver shall, without further action, succeed to the rights, titles and

the argument that the savings and loan association (S&L) had bargained away the rights of the FSLIC to bring a policy claim.[51] Rather, the issue was "whether public policy will allow [an S&L] to bargain away the rights of the FSLIC to carry out its statutory function."[52] The court's answer was a resounding "No." [53] Although the court acknowledged the importance of the freedom to contract, enforcing the exclusionary clause would "so seriously hamper the FSLIC in carrying out its duties that public policy prevents this court from enforcing the endorsement . . . ."[54]

A provision terminating coverage of an S&L's blanket bond upon takeover by a regulatory agency was at issue in *FSLIC v. Aetna Casualty & Surety Co.*[55] The provision was included in an agreement extending the discovery period for the S&L.[56] By terminating coverage when the FSLIC was appointed receiver, the agreement "granted less time to [the S&L's] successor, charged with protecting the public through the receivership and succeeding to all the rights and powers of [the S&L], to discover losses than [the S&L] had."[57]

Relying on various statutes and regulations[58] and *Oldenburg*,[59] the United States District Court for the Eastern District of Tennessee

---

privileges of the association. 12 C.F.R. § 547.7 (1987). The conservator shall immediately collect all obligations in money due the association and may . . . exercise all rights and powers of the association or execute, acknowledge and deliver any instrument necessary or proper for any purpose, and any instruments so executed shall be as valid and effectual as if it had been executed by the association's officers by authority of its Board of Directors. 12 C.F.R. § 548.2(b) and (h) (1987). The receiver shall collect all obligations and money due the association. The receiver may with respect to the association, exercise the powers which a conservator of a federal association may exercise under paragraphs (a) through (f) of section 548.2. 12 C.F.R. § 548.2(a) through (f) (1987).

Id. These provisions like all duly promulgated regulations have the full force and effect of law. *Id.* The enabling legislation for FSLIC, specifically § 1729, provides that when appointed as receiver the FSLIC is authorized to "*take over the assets of and operate such association*" and to "*collect all obligations to the insured institutions* . . . and to do *all* other things that may be necessary in connection therewith, *subject only to the regulation of the [FHLBB]* . . . . 12 U.S.C. § 1729 (1935) (emphasis added)." *Oldenburg*, 671 F. Supp. at 723.

51. *Id.*

52. *Id.*

53. *Id.* "Contractual provisions which are contrary to the terms and policy of a statute are illegal and unenforceable." *Id.* "This is especially true with respect to provisions in insurance policies." *Id.* at 724.

54. *Oldenburg*, 671 F. Supp. at 724.

55. 701 F. Supp. 1357 (E.D. Tenn. 1988).

56. *Id.* at 1358. "Such additional period of time shall terminate immediately . . . upon takeover of the Insured's business by any State or Federal official or agency, or by any receiver or liquidator . . . ." *Id.* (quoting the language of the bond at issue).

57. *Id.* at 1363.

58. 12 U.S.C. § 1729(d) (1989) (the FSLIC has the power to settle and

held the termination provision was void as against public policy because it restricted the exercise of the S&L's rights by the FSLIC. The contract attempted to "take away from the receiver rights that [the S&L] would have had against Aetna had there been no receivership. . . . This is clearly an infringement on the power of the FSLIC to act as a receiver of a failed institution, and this contract language is void and unenforceable."[60]

In *Branning v. CNA Insurance Co.*,[61] the policy at issue also contained a regulatory exclusion that precluded recovery for any claims brought by the FSLIC.[62] In refusing to enforce the clause, the United States District Court for the Western District of Washington found it "substantially hinders FSLIC's exercise of its federal powers and is therefore contrary to public policy."[63] Quoting 12 U.S.C. § 1729(d), which authorizes the FSLIC to terminate the affairs of insured institutions "*subject only to the regulation of the Federal Home Loan Bank Board*,"[64] and relying on *Oldenburg*,[65] the court held "[p]rivate parties to an insurance contract may not frustrate the Congressional purpose behind receivership by annulling FSLIC's federal powers."[66]

2. Issue Not Decided

The D&O insurance policy before the United States District Court for the Southern District of Ohio in *American Casualty Co. v. FSLIC* contained a "regulatory exclusion."[67] Claims "based upon or attributable to any action or proceeding *brought by or on behalf of* FSLIC" were excluded from coverage.[68] The S&L, however, filed suit against three officers ten days before it was declared insolvent and placed in receivership to the FSLIC.[69] When the officers notified

---

compromise bank claims); 12 C.F.R. § 547.7 (1990) (the receiver succeeds to all rights and powers of the defunct association); 12 C.F.R. § 549.3(a) (1990).

59. 671 F. Supp. 720, 723 (D. Utah 1987) (the FSLIC as receiver of a failed institution has all the powers and rights of that institution).

60. FSLIC v. Aetna Casualty & Sur. Co., 701 F. Supp. 1357, 1363 (E.D. Tenn. 1988).

61. 721 F. Supp. 1180 (W.D. Wash. 1989).

62. *Id.* at 1183-84.

63. *Id.* at 1184.

64. *Id.* (emphasis in original).

65. *Branning*, 721 F. Supp. at 1184.

66. *Id.* There is no indication the court considered *Continental Casualty*, decided only 16 days before *Branning*. *Continental Casualty* was decided on April 3, 1989. Continental Casualty Co. v. Allen, 710 F. Supp. 1088 (N.D. Tex. 1989). *Branning* was decided April 19, 1989. *Branning*, 721 F. Supp. at 1180.

67. 683 F. Supp. 1183 (S.D. Ohio 1988).

68. *Id.* at 1185 (emphasis in original).

69. *Id.* at 1184.

American Casualty they would seek indemnification of any liability under the policy, American filed a declaratory judgment action claiming that FSLIC, having substituted itself in place of the S&L in the D&O suit, eliminated any liability of American.[70] The FSLIC successfully argued the regulatory exclusion did not apply because the D&O suit was brought by the S&L, not "by or on behalf of" FSLIC.[71] Because of its decision that the action was not by or for the FSLIC, the court found it unnecessary to address the contention of the FSLIC that the regulatory exclusion violated public policy.[72]

3. Enforceable—Decision Not Based on Public Policy

A case involving another D&O bond was before the Court of Appeals for the Fifth Circuit in *Sharp v. FSLIC*.[73] The court noted that the FSLIC "requires all member banks to purchase insurance coverage under Savings and Loan Blanket Bond Standard Form No. 22 ("Form 22")."[74] Form 22 contained a provision that terminated coverage "immediately upon the taking over of the Insured by a receiver or other liquidator or by the State or Federal officials . . . ."[75] The Federal Home Loan Bank Board (FHLBB) appointed a conservator[76] to manage the S&L.[77] Although it was necessary to rule first on a notice provision of the bond,[78] the Fifth Circuit held coverage terminated when the conservator was appointed.[79] The bond did not cover losses discovered after the effective termination.[80]

In ruling the bond was terminated by the appointment of a conservator, the Fifth Circuit acknowledged the harshness of its decision but also articulated two factors that supported its conclusions. First, the FSLIC required coverage under Form 22. If it needed

---

70. *Id.* American Casualty argued FSLIC's substitution as plaintiff and removal of the D&O suit to a federal court constituted "bringing" the action in the district court. *Id.* at 1185.

71. *American Casualty*, 683 F. Supp. at 1185.

72. *Id.* at 1186.

73. 858 F.2d 1042 (5th Cir. 1988).

74. *Id.* at 1043.

75. *Id.* at 1044.

76. A conservator is an official charged with the protection of something affecting public welfare and interests. WEBSTER'S NEW COLLEGIATE DICTIONARY (1977).

77. *Sharp*, 858 F.2d at 1043.

78. *Id.* at 1044-46. The court rejected an argument that ambiguities in insurance contracts are to be strictly construed against the underwriter. *Id.* The court noted this rule is an extension of the general rule that contracts are to be construed strictly against the drafter, a rule not applicable to contracts that were a joint effort. *Id.* The court then stated Form 22 was a product of such a joint effort. *Id.* at 1046.

79. *Id.* at 1046.

80. *Id.* at 1044.

additional time to discover losses upon the appointment of a conservatorship, it could alter its regulations to require all S&Ls to purchase bonds that provide extended discovery time.[81] Second, the court reasoned, because the FSLIC and the FHLBB have the power to examine an S&L's books before institution of conservatorships or receiverships, and the losses that represent claims under blanket bonds are usually the same losses that lead the FHLBB to take over, the "sole effect of our decision is to require FSLIC and the FHLBB to do their homework prior to the institution of a conservatorship [or receivership]."[82] The *Sharp* court did not address the issue of voiding the exclusionary provision on public policy grounds.[83]

The United States District Court for the Northern District of Illinois had the opportunity to review a termination clause in a Form 22 blanket bond in *FSLIC v. Transamerica Insurance*.[84] While the court acknowledged the precedent established in *Oldenburg* and *FSLIC v. Aetna* for voiding the contract as against public policy,[85] it considered the determinative fact to be that the FSLIC requires S&L's to purchase insurance coverage under Form 22.[86] After quoting extensively from *Sharp*,[87] the court held the arguments of the FSLIC regarding the unenforceability of the termination provision must fail because "[h]aving required [the S&L] to procure a Form 22 bond, FSLIC simply cannot now be heard to complain about the enforceability of its provisions."[88]

4. Enforceable—No Violation of Public Policy

In *Continental Casualty Co. v. Allen*[89] the FDIC called upon the United States District Court for the Northern District of Texas[90] to void a clause in the D&O insurance policy that retracted coverage of the directors and officers for any action brought by or on behalf of the FDIC.[91] The court found that, unlike savings and loans regulated by FSLIC, D&O insurance is *optional* under all the FDIC and Office

---

81. *Sharp*, 858 F. Supp. at 1048.
82. *Id.*
83. 858 F.2d 1042 (5th Cir. 1988).
84. 705 F. Supp. 1328 (N.D. Ill. 1989). Section 11(c) provided the bond would be terminated "immediately upon the taking over of the Insured by a receiver or other liquidator or by State or Federal officials . . . ." *Id.* at 1332.
85. *Id.* at 1336.
86. *Id.*
87. *Transamerica Ins.*, 705 F. Supp. at 1337.
88. *Id.*
89. 710 F. Supp. 1088 (N.D. Tex. 1989).
90. The FDIC intervened in the suit between former directors and officers of a bank and the holder of the D&O insurance. The FDIC was heard in the case because of its interest in other lawsuits involving the bank's failure. *Id.* at 1097.
91. *Id.*

of the Comptroller of the Currency rules that applied to the Bank.[92] "Thus policies providing limited insurance, which are not required by statute or mandated as to form of coverage, are not invalidated on a public policy argument."[93]

Although the FDIC had cited several unpublished decisions[94] and *Oldenburg*[95] as authority for its public policy argument, the court found them unpersuasive.[96] The court said *Oldenburg* was based on another case that involved a statutorily required insurance minimum and inapplicable state law.[97] The unpublished decisions "are tenuously supported and often fail to provide adequate case law for the reasoning."[98] In addition, the court noted the unpublished decisions "sometimes rely on *Hudspeth*-type theories of deference to FSLIC or FDIC's receivership actions as rationale for refusing to enforce the endorsements."[99] The court took note of a very recent Supreme Court decision[100] that "reverses *Hudspeth*-type arguments and finds that a debtor is entitled to his day in court, despite the receivership. Thus, much of the 'public policy' deference and non-judicial interference with FDIC/FSLIC receiverships has been put to rest."[101]

In response to the FDIC's argument that to hold the endorsement enforceable gives banks the right to bargain away the FDIC's statutory right to take over and collect assets of a failed bank, the court opined that for the FDIC to collect an asset the failed bank must have had an asset.[102] In the court's view the failed bank

---

92. *Id.* at 1099 (emphasis added). "No FDIC-promulgated regulation is found regarding D&O insurance. Only a letter from the FDIC Chairman, not introduced as evidence in this case, indicates the FDIC's urging of banks not to accept insurance with such a FDIC endorsement on it. Such is not sufficient to make the . . . [insurance] policy in violation of law." *Id.* at 1099 n.20.

93. *Id.* at 1099.

94. The cases cited were: American Casualty Co. v. FSLIC, 704 F. Supp. 898 (E.D. Ark. 1989); FSLIC v. Mmahat, No. 86-5160, 1988 U.S. Dist. LEXIS 1079 (E.D. La. March 3, 1988); Maryland Deposit Ins. Fund Corp. v. American Casualty & Sur. Co., No. 88095087/CL 79669 (Md. Cir. Ct. Nov. 17, 1988).

95. 671 F. Supp. 720 (D. Utah 1987).

96. Continental Casualty Co. v. Allen, 710 F. Supp. 1088, 1099 (N.D. Tex. 1989).

97. *Id.* (citing Farmers Ins. Exch. v. Call, 712 P.2d 231 (Utah 1985)).

98. *Id. See generally* cases cited *supra* note 94.

99. *Id.* "*Hudspeth*-type theories of deference" is a reference to North Mississippi Sav. & Loan Ass'n v. Hudspeth, 756 F.2d 1096 (5th Cir. 1985), *cert. denied*, 474 U.S. 1054 (1986), which held "a claim against the FSLIC must first be pursued administratively so that the courts do not interfere with the receivership." *Id.* at 1099 n.22.

100. Coit Independence Joint Venture v. FSLIC, 489 U.S. 561 (1989). The Court held the statutes governing the FSLIC and FHLBB do not grant the FSLIC adjudicatory power over the claims of creditors of insolvent S&Ls under FSLIC receivership and do not divest the courts of jurisdiction to consider these claims *de novo*. *Id.* at 1368.

101. 710 F. Supp. at 1099.

102. *Id.* at 1099-1100.

> did not own as an asset directors' and officers' liability insurance without an FDIC endorsement. Therefore, such argument fails. Nothing in the endorsement affects FDIC as receiver to have all the rights and claims that the failed banking institution would have had. . . . FDIC steps into the shoes of the failed bank and is subject to whatever contracts the bank, under its freedom of contract, entered.[103]

B. Most Recent Court of Appeals Case—FDIC v. Aetna

The Court of Appeals for the Sixth Circuit considered the enforceability of a regulatory exclusion in the 1990 case *FDIC v. Aetna*.[104] Two blanket bonds issued by Aetna contained identical sections providing they would terminate upon the takeover of the insured by the FDIC, and the FDIC could not purchase an additional time period to discover loss.[105] The trial court, the United States District Court for the Middle District of Tennessee, had ruled the exclusionary provisions void as against public policy and entered an award for damages to the FDIC.[106] The Sixth Circuit overturned the lower court's determination and held public policy was not violated by regulatory exclusion provisions where the policies are not required by statute and where the form of coverage has not been mandated.[107] Three interrelated issues formed the basis of the court's decision: (1) the power of courts generally in connection with public policy and contracts; (2) legislative action regarding regulatory exclusions in financial institution D&O policies; and (3) FDIC regulatory requirements regarding D&O insurance.[108]

1. Public Policy

Having noted when a district court construes a contract the interpretation is a question of law and reviewable de novo by the appellate court,[109] the Sixth Circuit reviewed "well-settled principles" to determine whether the contracts were contrary to public policy.[110] One excerpt from *Muschany* cited by the court said "[i]t is a matter

---

103. *Id.* at 1100.
104. 903 F.2d 1073 (6th Cir. 1990).
105. *Id.* at 1074-75.
106. *Id.* at 1077. The lower court found the regulatory exclusions contrary to public policy because they "preclude the FDIC from discharging its responsibility in connection with marshalling the assets of the failed bank." *Id.* The court also stated that to enforce the exclusions would be to sanction "the bargaining away of FDIC's statutory function upon being appointed as receiver." *Id.*
107. *Id.* at 1078 (citing *Continental Casualty*, 710 F. Supp. at 1099).
108. *FDIC v. Aetna*, 903 F.2d at 1077-79.
109. *Id.* at 1078.
110. *Id.*

of public importance that good faith contracts of the United States should not be lightly invalidated. Only dominant public policy would justify such action." However, while *Muschany* referred to contracts between the United States government (specifically the War Department) and private citizens, the Sixth Circuit was considering a contract between two private parties.[111] Although the Supreme Court has stated elsewhere contractual obligations undertaken by private citizens should seldom be undermined on public policy grounds,[112] the *Muschany* Court clearly appeared to be influenced by the fact that the party seeking to avoid its obligations was the United States government. The policies underlying the *Muschany* decision may not be present to the same degree in cases involving only private citizens. Thus, the Sixth Circuit's wholesale reliance on *Muschany* may be misplaced.

The Sixth Circuit also relied on the opinion of the Fourth Circuit in *St. Paul Mercury Insurance Co. v. Duke University*[113] when it reached the conclusion that "questions of public policy are to be determined in the *first* instance by the legislature."[114] While the Supreme Court has said "[p]rimarily it is for the lawmakers to determine the public policy,"[115] the Sixth Circuit ignored the Court's acknowledgement that reference to other sources when considering public policy is also permissible. The Supreme Court has stated "[i]n determining whether [a] contract . . . contravenes the public policy of [a state], the Constitution, laws, and *judicial* decisions of that state, as well as the applicable principles of the common law, are to be considered."[116] It would appear not only permissible but inescapable to refer to other sources when the legislature has not yet made a determination on a specific issue, such as the enforceability of regulatory exclusions.

2. Legislative Action Regarding Public Policy

The Sixth Circuit based its determination that regulatory exclusions did not violate public policy primarily on the 1989 Amendment of 12 U.S.C. § 1821 which states, in § 1821(e)(12)(A), a receiver "may enforce any contract *other than a director's or officer's liability insurance contract or a depository institution bond*, entered into by the depository institution notwithstanding any provision of the contract providing for the termination . . . upon, or solely because of,

---

111. Muschany v. United States, 324 U.S. 49, 66 (1945).

112. Twin City Pipe Line Co. v. Harding Glass Co., 283 U.S. 353 (1931).

113. 849 F.2d 133 (4th Cir. 1988).

114. FDIC v. Aetna Casualty & Sur. Co., 903 F.2d 1073, 1078 (6th Cir. 1990) (emphasis added).

115. *Id.*

116. *Twin City*, 283 U.S. at 357 (emphasis added).

insolvency or the appointment of a . . . receiver."[117] The court held this statute "[did] not provide the basis for a 'dominant public policy' which would justify voiding [the regulatory exclusions]."[118]

While the language of § 1821(e)(12)(A) suggests Congress was taking a position on the issue, there was evidence directly to the contrary that the Sixth Circuit completely ignored. Title 12 U.S.C. § 1821(e)(12)(B), speaking of § 1821(e)(12) as a whole, states: "No provision of this paragraph may be construed as impairing or affecting any right of the conservator or receiver to enforce or recover under a directors or officers liability insurance contract or depository institution bond under other applicable law." In addition, the Sixth Circuit neglected to consider the legislative history, which clearly indicates Congress was *not* trying to settle this area of the law, an area repeatedly litigated with conflicting results,[119] when § 1821 was passed.[120] House Report 54 clearly states the 1989 amendment is neutral:

> The bill as reported by the Committee retains current law for the treatment of exclusionary clauses in directors and officers liability insurance contracts or financial institution bonds. Nothing in [the amended statute] impairs or affects any rights a conservator or receiver already appointed [or appointed in the future] had under current law . . . to enforce or recover under a directors or officers insurance policy contract or financial institution bond.
>
> . . . .
>
> . . . The exception for directors and officers liability contracts and financial institution bonds [in 12 U.S.C. § 1821(e)(12)(A)] is intended neither to expand nor diminish the rights of insured institutions, their insurers, or conservators or receivers. A conservator or receiver or insurer retains the right to litigate the enforceability of coverage exclusions . . . that state that upon appointment of a receiver coverage will be terminated. Courts shall continue to

---

117. 903 F.2d at 1078 (emphasis in original).

118. *Id.*

119. Sharp v. FSLIC, 858 F.2d 1042 (5th Cir. 1988) (coverage terminates upon takeover by FSLIC); Branning v. CNA Ins. Co., 721 F. Supp. 1180, 1184 (W.D. Wash. 1989) (regulatory exclusion void as against public policy); Continental Casualty Co. v. Allen, 710 F. Supp. 1088, 1099 (N.D. Tex. 1989) (regulatory endorsement not void on public policy grounds); FSLIC v. Transamerica Ins., 705 F. Supp. 1328 (N.D. Ill. 1989) (termination provision enforced); FSLIC v. Aetna Casualty & Sur. Co., 701 F. Supp. 1357 (E.D. Tenn. 1988) (termination provision void as against public policy); FSLIC v. Oldenburg, 671 F. Supp. 720 (D. Utah 1987) (exclusionary endorsement violates public policy). USBN was declared insolvent in 1983 and the FDIC filed its complaint in 1985. FDIC v. Aetna Casualty & Sur. Co., 903 F.2d at 1075-76. At all crucial times before the court in *FDIC v. Aetna*, Congress had not enacted any law addressing the subject.

120. H.R. REP. No. 54, 101st Cong., 1st Sess., pt. 1, *reprinted in* 1989 U.S.C.C.A.N. 86, 127, 212-13.

> address provisions in such contracts and bonds under existing law. This provision is in no way intended to affect or alter the law with respect to directors and officers liability insurance contracts or financial institution bonds. . . .
>
> The majority of courts which have considered the provisions in contracts which provide that coverage terminates upon appointment of a receiver have found such provisions to be against public policy and therefore unenforceable. In a recent case, the Court held that such a clause "substantially hinders FSLIC's exercise of its federal powers and therefore is contrary to federal policy." The Court stated further: "If the court were to enforce the FSLIC exclusion as written, all of FSLIC's claims, regardless of their origin or status under the policy, would not be covered simply because FSLIC rather than a shareholder, depositor, or third party prosecuted the claim. *Private parties to an insurance contract may not frustrate the Congressional purpose behind receivership by annulling FSLIC's federal powers.*" *Branning v. CNA Insurance Companies*, [721 F. Supp. 1180, 1184 (W.D. Wash. 1989) (emphasis in original)]. For a contrary view, see *Continental Casualty Co. v. Allen*, [710 F. Supp. 1088 (N.D. Tex. 1989)].[121]

The Sixth Circuit's determination in *FDIC v. Aetna* that "[t]hese statutes do not provide the basis for a 'dominant public policy' which would justify voiding [the regulatory exclusions]" is arguably correct when only the specific statutory language of § 1821(e)(12)(A) is considered. The foregoing discussion, however, demonstrates § 1821(e)(12)(A) cannot be considered in isolation. While the Sixth Circuit was within the bounds of the law when it reached the decision it did, its analysis was based on one statutory provision taken out of context so as to be very misleading, with no consideration of the legislative history that specifically addressed the issue.[122] Given congressional acknowledgment that a majority of prior cases found a violation of public policy and a clear statement that courts are to continue developing this area of the law, the Sixth Circuit's incomplete consideration of the issue does not appear consistent with congressional intent. Unfortunately, other courts have blindly adopted the Sixth Circuit's ruling.[123] If this trend continues, the courts' incomplete analysis will forever subvert the intent of Congress that the courts continue to develop this area of the law.

---

121. *Id.*

122. 903 F.2d at 1078.

123. American Casualty Co. v. Baker, 758 F. Supp. 1340, 1345-47 (C.D. Cal. 1991); St. Paul Fire & Marine Ins. Co. v. FDIC, 765 F. Supp. 538, 549-50 (D. Minn. 1991) (the language in these cases closely tracks the language of *FDIC v. Aetna*).

3. Regulatory Requirements for D&O Insurance

Another factor crucial to the courts' decision in *FDIC v. Aetna* that the exclusionary clause does not violate public policy was that the FDIC does not require member institutions to obtain D&O insurance.[124] Title 12 U.S.C. § 1828(e)[125] provides that the FDIC may require an insured bank to purchase fidelity bond coverage. However, no evidence was presented indicating the FDIC had exercised its authority.[126]

In discussing the importance of this consideration, the *Aetna* court relied heavily on *Sharp v. FSLIC*,[127] which found "no public policy against a termination provision identical to [the one] in the instant case."[128] However, *Sharp* may be distinguished because one of the dispositive issues for the Fifth Circuit was that the FSLIC required all member S&Ls to purchase Form 22 insurance coverage and Form 22 contains the "termination upon takeover" clause.[129] They enforced the terms of the contract because the FSLIC specifically required coverage under Form 22.[130] Additionally, the *Sharp* court did not address any public policy argument.[131] Its decision could be read as having considered and rejected a public policy argument, but it does not specifically state its opinion regarding public policy.[132]

The Sixth Circuit approvingly cited the contention of other courts that to avoid the problem regarding termination provisions the FDIC and FSLIC could: (1) require bonds that continue upon appointment of a receiver;[133] and (2) "do their homework before institution of a receivership," which would presumably lead to the filing of claims

---

124. A district court cited this determination to bolster its own later conclusion that this is a significant point. Gary v. American Casualty Co., 753 F. Supp. 1547 (W.D. Okla. 1990). Having concluded regulatory exclusions do not violate public policy where such policies are not required by statute, the Sixth Circuit proceeded to dismiss the FDIC's argument that the parties to the insurance contract have unequal bargaining power because Tennessee law requires the purchase of the bonds. 903 F.2d at 1078.

125. 12 U.S.C. § 1828(e) (1988).

126. 903 F.2d at 1078. This point was also considered, with similar results, in *Continental Casualty*. There, the court stated "[n]o FDIC-promulgated regulation is found regarding D&O insurance. Only a letter from the FDIC chairman, not introduced as evidence in this case, indicates the FDIC's urging of banks not to accept insurance with such an FDIC endorsement on it." *Continental Casualty*, 710 F. Supp. at 1099 n.20.

127. 858 F.2d 1042 (5th Cir. 1988).

128. 903 F.2d at 1078.

129. 858 F.2d 1042 (5th Cir. 1988).

130. *Id.*

131. *Id.*

132. *See* 903 F.2d 1073 (6th Cir. 1989).

133. *Id.* at 1078 (citing *Continental Casualty*, 710 F. Supp. at 1099).

prior to a receiver being appointed.[134] The only support for the FDIC and FSLIC not implementing the first option would be if the market for D&O coverage is so tight that such a requirement would make it impossible to obtain. There is evidence that would lead to the conclusion this is indeed the reason for not requiring coverage that remains in effect for a receivership.[135] State chartered institutions are under the supervision of state commissioners of banking, and federally chartered institutions are under the supervision of the Office of the Comptroller of the Currency. Therefore, option number two, in theory, and perhaps in reality, is beyond the control of the regulatory agencies.

In addition, the court cited *Sharp* for the proposition that refusing to enforce regulatory terminations or exclusions would allow the FDIC to force insurers to take risks they did not bargain for because "the officials who purchased the bond to insure their own honesty are no longer in control of the institution. Takeover or receivership would substantially alter the character of the risk covered by the policy."[136] This criticism would be valid if the FDIC were trying to require insurers to provide coverage for events occurring after a takeover or receivership, but that is not the case. The FDIC is attempting to force insurers to cover claims for actions that occurred prior to a bank's closing but were not discovered until after the institution was taken over or closed. That is by nature the function of "occurrence policies," and is no greater risk than bargained for in "claims made" policies during their effective dates.[137]

The Sixth Circuit determined that "the FDIC's contention that existing laws justify the invalidation of the regulatory exclusion must fail. The dominant public policy exposed by this review is that the parties' freedom of contract must not be disturbed."[138] However, in light of the above arguments, this conclusion is questionable.

IV. CONCLUSION

It may be true that the liberty to contract freely is the dominant public policy, as expressed in *Muschany*. Courts, however, have

---

134. *Id.* (citing *Transamerica Ins.*, 705 F. Supp. at 1328).

135. In a speech to the American Bankers Association in 1986, the FDIC Chairman discussed efforts the agency is pursuing to assist banks in obtaining D&O insurance. Included were discussions with insurers and reinsurers in the United States and London, as well as working with other regulators to develop guidelines for directors and officers. JOURNAL OF ACCOUNTANCY 50 (Jan. 1987).

136. *Id.* (citing *Continental Casualty*, 710 F. Supp. at 1088; *Transamerica Ins.*, 705 F. Supp. at 1328).

137. There are two types of D&O policies—"claims made" policies cover claims made during the time period covered by the policy; "occurrence" policies cover claims arising from events that occurred during the policy period, regardless of when the claims were actually made. Branning v. CNA Ins. Co., 721 F. Supp. 1180, 1183 (W.D. Wash. 1989).

138. 903 F.2d 1073 (6th Cir. 1989).

clearly expressed a willingness to void contracts as against public policy in appropriate circumstances, and the Sixth Circuit failed to adequately consider indications that cases involving regulatory exclusion clauses may present such circumstances.

In *FDIC v. Aetna*, the Sixth Circuit declined to follow *Branning v. CNA*,[139] *FSLIC v. Aetna*,[140] and *FSLIC v. Oldenburg*,[141] even in the face of an expression that Congress was satisfied with the result reached in those cases.[142] The analysis put forth in this effort is unpersuasive and represents a selective recitation of the law in an attempt to establish a public policy for termination and regulatory exclusion endorsements in D&O policies and blanket bonds. Under the guise of rigourous enforcement of contractual obligations, the court engaged in a selective recitation of the law in an attempt to reach a desired result in spite of both law and fact.

Therefore, the courts have yet to satisfactorily address the issue of the validity of regulatory exclusions. Congress has taken a neutral position. Some of the early cases did not deal with the public policy issue, and the most recent case decided by a federal court of appeals, *FDIC v. Aetna*, is inadequately supported. A consensus cannot be reached by the courts that have addressed the issue to date. The only foreseeable resolution is if the Supreme Court agrees to hear one of the cases or Congress takes a position as to the public policy issue.

In the meantime, insurance companies can deny coverage if the claim was not filed before the financial institution failed. These unrecoverable claims further deplete the regulatory deposit insurance funds. This increases the cost of bailouts to taxpayers for insolvent regulatory funds and the deposit insurance premiums paid by member institutions of solvent funds.

LINDA C. ELAM

139. 721 F. Supp. 1180 (W.D. Wash. 1989).
140. 701 F. Supp. 1357 (E.D. Tenn. 1988).
141. 671 F. Supp. 720 (D. Utah 1987).
142. *See supra* note 120.

TENNESSEE'S LONG-AWAITED ADOPTION OF PROMISSORY FRAUD: *Steed Realty v. Oveisi**

> And be these juggling fiends no more believed, That palter with us in a double sense; That keep the word of promise to our ear And break it to our hope.[1]

I. INTRODUCTION

On September 3, 1991, the Tennessee Supreme Court denied permission to appeal in *Steed Realty v. Oveisi*.[2] The Tennessee Supreme Court's denial of permission to appeal in *Steed Realty* represents the culmination of a lengthy development of the law in Tennessee regarding the doctrine of promissory fraud.[3] In affirming

---

* I want to thank J.E. Gervin, Jr. for introducing me to the issue of promissory fraud, Donald J. Aho, Benjamin Y. Pitts, and Professor John Sobieski for their helpful comments, and Professor Carol Mutter for her suggested revisions to the initial draft of this Comment.

1. WILLIAM SHAKESPEARE, MACBETH act 5, sc. 8.

2. 1991 WL 288197 (Tenn. Ct. App. May 8, 1991), *appeal denied*, Sept. 3, 1991.

3. The full impact that *Steed Realty v. Oveisi* will have on the state of the law in Tennessee regarding promissory fraud remains to be seen. As of the date of this writing, *Steed Realty* has not been published. It is this writer's opinion, however, that it will be published. Rule 11 of the Rules of the Court of Appeals of Tennessee sets forth the criteria for determining the publication of opinions where no application for permission to appeal to the Tennessee Supreme Court is filed. If equal consideration is given, which arguably it should be, to the determination of publication of opinions where an application for permission to appeal is filed and denied by the Tennessee Supreme Court, the criteria would be the same as those listed in Rule 11. Rule 11 of the Rules of the Court of Appeals of Tennessee states:

> (b) An opinion of the Court . . . shall be published only if, in the determination of the members of the Court, it meets one or more of the following criteria:
> (1) the opinion establishes a new rule of law or alters or modifies an existing rule or applies an existing rule to a set of facts significantly different from those stated in other published opinions;
> (2) the opinion involves a legal issue of continuing public interest;
> (3) the opinion criticizes, with reasons given, an existing rule of law;
> (4) the opinion resolves an apparent conflict of authority;
> (5) the opinion updates, clarifies or distinguishes a principle of law; or
> (6) the opinion makes a significant contribution to legal literature by reviewing . . . the development of a common law rule.

Assuming the opinion is published, a question then arises as to its precedential effect, that being, what precedential value is to be given to the denial by the supreme

the trial court, the Tennessee Court of Appeals found the doctrine of promissory fraud applicable to the facts before it[4] and went on to grant rescission of contracts for the purchase of real estate on that basis.[5] *Steed Realty* is the case for which the legal community in Tennessee has been waiting since 1967 when the Tennessee Supreme Court, in *Bolan v. Caballero*,[6] announced its willingness to recognize the doctrine of promissory fraud "in a proper case where justice demands . . . ."[7] This comment will examine (1) the doctrine of

---

court of an application for permission to appeal? Tennessee Supreme Court Rule 4 indicates that there are two types of denials of permission to appeal: (1) a denial concurring in result only (commonly referred to as a "DCRO"), and (2) a denial without restricting language. To resolve uncertainty surrounding the precedential value of each of these denials, the Tennessee Supreme Court Commission to Study Appellate Courts currently is holding hearings and soliciting comments on this matter. *See* Request for Comments on Improving Appellate Justice in Tennessee, 812 S.W.2d No. 4, CXII (1991) [hereinafter Request for Comments]. Despite any apparent uncertainty, some authority holds that when the supreme court denies review concurring in result only, the supreme court affirms the result but not necessarily the reasoning of the intermediate appellate court. *See* Request for Comments, *supra*; Lawrence A. Pivnick & James C. Schaeffer, TENNESSEE CIRCUIT COURT PRACTICE § 30-12 (3d ed. 1991) [hereinafter Pivnick]; Adams v. State, 547 S.W.2d 553, 556 (Tenn. 1977); Clingan v. Vulcan Life Ins. Co., 694 S.W.2d 327, 331 (Tenn. Ct. App. 1985).

In contrast, when the supreme court denies permission to appeal without limitation, this disposition amounts to the supreme court's approval of both the result and the reasoning of the intermediate appellate court and is entitled to stare decisis effect. Request for Comments, *supra*; Pivnick, *supra*; Hathaway v. Middle Tennessee Anesthesiology, P.C., 724 S.W.2d 355, 366 n.1 (Tenn. Ct. App. 1986). *See also* Spalding v. Davis, 674 S.W.2d 710 (Tenn. 1984) (supreme court's denial of permission to appeal that could not be interpreted as a denial concurring in result only "overruled" a prior supreme court decision and was binding on the case at bar); Pairamore v. Pairamore, 547 S.W.2d 545, 548 (Tenn. 1977) ("When we deny the writ [for certiorari] we . . . have made a final disposition of the case by approving the final decree of the [c]ourt of [a]ppeals."); Beard v. Beard, 14 S.W.2d 745, 747 (Tenn. 1929) ("Denial of the writ of certiorari to review the action of the [c]ourt of [a]ppeals, without a written opinion or some explanatory memorandum, emphasizes the concurrence of the [c]ourt in the opinion of the [c]ourt of [a]ppeals."); Pankow v. Mitchell, 737 S.W.2d 293, 297 n.3 (Tenn. Ct. App. 1987) ("The Tennessee Supreme Court denied . . . applications for permission to appeal in these cases, thereby indicating its approval in the reasoning and the results of the two opinions.").

In *Steed Realty* the Tennessee Supreme Court did not indicate that it denied permission to appeal concurring in result only. Therefore, because the denial was without limitation, the Tennessee Supreme Court apparently approved of both the reasoning and result reached by the appellate court. This disposition by the supreme court in *Steed Realty*, then, appears to be Tennessee's adoption of the doctrine of promissory fraud.

4. *Steed Realty*, 1991 WL 288197 at *6.

5. *Id.*

6. 417 S.W.2d 538 (Tenn. 1967). *See infra* notes 120-33 and accompanying text.

7. *Id.* at 541.

promissory fraud,[8] (2) Tennessee's position regarding promissory fraud,[9] (3) the *Steed Realty* decision,[10] and (4) the effects of *Steed Realty*.[11]

II. THE DOCTRINE OF PROMISSORY FRAUD

Ordinarily a prediction that events will occur in the future is to be regarded as opinion only, on which a party has no right to rely.[12] Mere opinion is generally nonactionable in fraud[13] because the inherent vagueness and lack of certainty of opinions should forestall a plaintiff's reliance.[14] As has often been stated, a man "'can not warrant a thing which will happen in the future.'"[15] As the court in *Scott v. United States*[16] stated: "Mere puffing, exaggerated enthusiasm and high pressure salesmanship does not constitute legal fraud. This is also true as to unfulfilled promises, prophecies, predictions and erroneous conjecture as to future events . . . ."[17]

An important distinction exists, however, between predictions of events to occur in the future when those events are outside the speaker's control and statements about what the speaker himself will do in the future.[18] In the latter case, such statements of intention are usually regarded as statements of fact on which a party's reliance is justified.[19]

Historically the law recognized that every statement, whether expressing a mere opinion or a prediction of events to occur in the

---

8. *See infra* notes 12-89 and accompanying text.
9. *See infra* notes 90-223 and accompanying text.
10. *See infra* notes 224-51 and accompanying text.
11. *See infra* notes 252-66 and accompanying text.
12. W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 109 at 762 (5th ed. 1984) [hereinafter PROSSER & KEETON].
13. As one court stated, the six elements of the tort of fraud in Tennessee are: (1) "a representation of an existing or past fact and not an opinion or a conjecture as to future events," (2) "the representation must be false," (3) "the representation must be in regard to a material fact," (4) "proof of fraud," (5) "the plaintiff must rely reasonably on the misrepresented material fact," and (6) "the plaintiff must suffer damage." Edwards v. Travelers Ins., 563 F.2d 105, 110-13 (6th Cir. 1977). *See* Haynes v. Cumberland Builders, Inc., 546 S.W.2d 228, 232 (Tenn. Ct. App. 1976).
14. Michael J. Polelle, *An Illinois Choice: Fossil Law or an Action for Promissory Fraud?*, 32 DEPAUL L. REV. 565, 566 (1983) [hereinafter Polelle].
15. Fleming James & Oscar Gray, *Misrepresentation—Part II*, 37 MD. L. REV. 488, 502 (1978) [hereinafter James & Gray] (quoting Choke, J., in Anonymous, Y.B. Pasch. 11 Edw. 4, f.6, pl. 11 (1471)).
16. 263 F.2d 398 (5th Cir. 1959).
17. *Id.* at 401 n.2 (*quoting* Defendant's Requested Instruction No. 10) (citations omitted).
18. James & Gray, *supra* note 15, at 502.
19. PROSSER & KEETON, *supra* note 12, at 764.

future, contained some degree of factual assertion.[20] Similarly, with regard to promises or statements of intention as to events within the speaker's control, the courts found when a promisor made his promise he represented as factual his present intent to take some action in the future to fulfill his promise.[21] Thus, when a promisor made a promise with no present intent to fulfill it, he misrepresented an existing fact—his present state of mind.[22] As a result, the promisee who relied on this factual misrepresentation was found to have a sufficient basis for an action of fraud.[23]

The doctrine of promissory fraud, that is, promissory statements made with the present intent not to fulfill them,[24] originated in the case of *Edington v. Fitzmaurice*.[25] In *Edington* a group of investors purchased corporate debentures in reliance on the representations of the company's directors that the money from the sale of the debentures would be used to purchase new equipment and to improve the company's facilities.[26] The directors, however, had no intention of improving the corporate facilities.[27] Rather, they intended to use, and, in fact, did use the money to pay off existing liabilities.[28] Holding the directors liable for fraud, Lord Bowen found:

> The state of a man's mind is as much a fact as the state of his digestion. It is true that it is very difficult to prove what the state of a man's mind at a particular time is, but if it can be ascertained it is as much a fact as anything else. A misrepresentation as to the state of a man's mind is, therefore, a misstatement of fact.[29]

---

20. Polelle, *supra* note 14, at 566.

21. *Id.*; PROSSER & KEETON, *supra* note 12, at 763; James & Gray, *supra* note 15, at 506; W. Page Keeton, *Fraud: Misrepresentations of Opinion*, 21 MINN. L. REV. 643, 644 (1937); W. Page Keeton, *Fraud—Statements of Intention*, 15 TEX. L. REV. 185, 186 (1937).

22. PROSSER & KEETON, *supra* note 12, at 763; James & Gray, *supra* note 15, at 506. *See also* Elk Ref. Co. v. Daniel, 199 F.2d 479, 481 (4th Cir. 1952) ("the presence or lack of a positive intention at any time is a factor of a state of mind, and a misstatement of that mental position is a false representation of an existing fact.").

23. PROSSER & KEETON, *supra* note 12, at 763; Polelle, *supra* note 14, at 566; Note, *The Legal Effect of Promises Made With Intent Not to Perform*, 38 COLUM. L. REV. 1461 (1938). *See also* Street v. J.C. Bradford & Co., 886 F.2d 1472, 1483 n.30 (6th Cir. 1989) (misrepresentations as to future events can be the basis of fraud if there is no present intent or ability to perform them).

24. For discussion of the definition of promissory fraud, see Fowler v. Happy Goodman Family, 575 S.W.2d 496, 499 (Tenn. 1978) (discussed *infra* at notes 134-47 and accompanying text).

25. 29 Ch. D. 459 (1885).

26. *Id.* at 460-61.

27. *Id.* at 461-62.

28. *Id.* at 462.

29. *Id.* at 483.

Lord Bowen's conclusion that a misrepresentation of one's state of mind is a misstatement of fact established the doctrine of promissory fraud as an action in tort separate and distinct from an action on the contract.[30] This is a fine distinction because both a contract and a tort action may arise out of the same set of facts, but the mere breach of a promise or failure to perform one's obligations under a contract does not provide the basis for an action for promissory fraud.[31] A promise may be left unfulfilled because the promisor subsequently finds himself unable to fulfill it, because future events did not occur as originally predicted,[32] or for any number of reasons. In these examples there is no misrepresentation unless the promisor's state of mind at the time he makes the promise is misstated.[33] Put another way, no cause of action for fraud arises when a promise made in good faith with the expectation of being performed is subsequently broken.[34] Otherwise, any breach of contract would give rise to an action for fraud.[35]

This contract-tort distinction is again important because it determines the treatment of two policies critical to the doctrine of promissory fraud: (1) the Statute of Frauds[36] and (2) the Parol Evidence Rule,[37] both discussed below.

A. The Statute of Frauds

The success of an action for promissory fraud will depend significantly on the classification of the action as either tort or contract. If the action is deemed to be one on the contract to which the Statute of Frauds applies, the plaintiff's promissory fraud action will be seriously impaired.[38] Basically, the purpose of the Statute of Frauds

---

30. Polelle, *supra* note 14, at 568. *See also* PROSSER & KEETON, *supra* note 12, at 763.

> All but a few courts regard a mistatement of a present intention as a misrepresentation of a material fact; and a promise made without the intent to perform it is held to be a sufficient basis for an action of deceit, or for restitution or other equitable relief. . . . The door is thus opened to a tort remedy which may offer important advantages over any action on the contract itself . . . .

Id. (citations omitted).

31. PROSSER & KEETON, *supra* note 12, at 764; Polelle, *supra* note 14, at 567.

32. Polelle, *supra* note 14, at 567-68.

33. PROSSER & KEETON, *supra* note 12, at 764.

34. *Id.*

35. *Id.*

36. *See infra* notes 38-75 and accompanying text.

37. *See infra* notes 76-89 and accompanying text.

38. *See* Evan M. Zuckerman, Note, *Promissory Fraud in Tennessee: A Wrong without a Remedy*, 10 MEM. ST. U.L. REV. 308 (1980) [hereinafter Note, *Promissory*

is to prevent fraudulent contracts from being proved by perjured testimony.[39] This purpose is accomplished under the Statute of Frauds by the requirement that promises within the Statute be in writing.[40] Where a plaintiff-promisee bases his claim on an oral promise allegedly made to him by the promisor, enforcement of the promise conflicts with the Statute of Frauds. Thus, if the policy behind the Statute of Frauds is to be honored, that is, preventing the establishment of a fraudulent contract by perjured testimony,[41] a promissory fraud action must be disallowed by the Statute of Frauds.

An equally important and compelling policy behind the Statute of Frauds, however, is to prevent the promisor, after having not performed his oral promise, from hiding behind the Statute of Frauds and using it as a shield to protect him in his fraud.[42] Just as the Statute of Frauds prevents a dishonest plaintiff from using false testimony to invent a nonexistent contract based on alleged oral promises, it also prevents a truly defrauded plaintiff from establishing the existence of an oral contract that was in fact made.[43] The Statute's potential for causing injustice is its greatest in this latter situation, where the Statute of Frauds is pleaded as a defense to an action in tort for fraud based on an oral promise within the Statute.[44]

It is in this context, then, that the determination as to the applicability of the Statute of Frauds to an action for promissory

---

Fraud in Tennessee]; Polelle, *supra* note 14.

The Statute of Frauds, at TENN. CODE ANN. § 29-2-101(a) (Supp. 1991), states:

> Writing required for action.—(a) No action shall be brought:
> (1) Whereby to charge any executor or administrator upon any special promise to answer any debt or damages out of his own estate;
> (2) Whereby to charge the defendant upon any special promise to answer for the debt, default, or miscarriage of another person;
> (3) Whereby to charge any person upon any agreement made upon consideration of marriage;
> (4) Upon any contract for the sale of lands, tenements, or hereditaments, or the making of any lease thereof for a longer term than one (1) year; or
> (5) Upon any agreement or contract which is not to be performed within the space of one (1) year from the making thereof;
> unless the promise or agreement, upon which such action shall be brought, or some memorandum or note thereof, shall be in writing, and signed by the party to be charged therewith, or some other person by him thereunto lawfully authorized.

39. Baliles v. Cities Serv. Co., 578 S.W.2d 621, 623 (Tenn. 1979); Yates v. Skaggs, 213 S.W.2d 41 (Tenn. 1948).

40. *See* TENN. CODE ANN. § 29-2-102 (Supp. 1991).

41. *See supra* note 39 and accompanying text.

42. *See Baliles*, 578 S.W.2d at 623 ("The purpose of the statute of frauds is to reduce contracts to a certainty, in order to avoid perjury on the one hand and fraud on the other."), *quoting* Price v. Tennessee Prods. & Chem. Corp., 385 S.W.2d 301 (Tenn. Ct. App. 1964).

43. George N. Stepaniuk, Note, *The Statute of Frauds as a Bar to an Action in Tort for Fraud*, 53 FORDHAM L. REV. 1231, 1232 (1985).

44. *Id.* at 1232-33.

fraud must be made. This determination necessarily results in a balancing of policy considerations as to whether the interests of justice are best served by (1) disallowing an action for promissory fraud because doing so apparently contravenes the policy behind the Statute of Frauds, or by (2) allowing an action for promissory fraud because to disallow it would protect and even reward the dishonest promisor via an abuse of the Statute of Frauds.[45] The true issue, then, is whether the Statute of Frauds operates as a bar to a tort claim based on a fraudulent oral promise that is rendered unenforceable as a contract by the Statute.[46]

Helpful to the determination of the applicability of the Statute of Frauds is the effect of Lord Bowen's famous pronouncement in *Edington* that the state of a man's mind is as much a fact as the state of his digestion:[47] it established the action of promissory fraud as an action in tort.[48] As the rule in Tennessee is that the Statute of Frauds does not apply to actions in tort,[49] the Statute appears to be inapplicable to bar a plaintiff's promissory fraud claim.

Applying Tennessee law, the Court of Appeals for the Sixth Circuit explained the inapplicability of the Statute of Frauds to a tort action in *Jarrett v. Epperly*.[50] In that case defendant Epperly approached Jarrett and asked him to come work for him.[51] Only after Epperly promised Jarrett forty-nine percent ownership of the business after ten years of service did Jarrett accept the offer.[52] Just before his tenth anniversary with the company, Jarrett requested from Epperly performance of the oral promise.[53] Epperly acknowledged the promise and told Jarrett he would receive proof of his forty-nine percent ownership.[54] Shortly thereafter, however, Epperly sold the assets of the company for over $11 million,[55] never honoring his oral promise. Jarrett filed suit seeking damages for breach of an oral employment contract and for promissory fraud.[56]

---

45. *See id.* at 1234 ("The split in authority on this issue thus reflects a disagreement as to which type of fraud—that of the promisor or the promisee—poses the greater threat to the Statute's overall policy of preventing fraud in business dealings.").

46. *Id.* at 1233.

47. *Supra* text accompanying note 29.

48. *Supra* note 30 and accompanying text. *See also* PROSSER & KEETON, *supra* note 12, at 764 ("The tendency is clearly to treat the misrepresentation as a separate matter from the contract.").

49. Jarrett v. Epperly, 896 F.2d 1013 (6th Cir. 1990) (applying Tennessee law); Haynes v. Cumberland Builders, Inc., 546 S.W.2d 228 (Tenn. Ct. App. 1976).

50. 896 F.2d 1013 (6th Cir. 1990).

51. *Id.* at 1015.

52. *Id.*

53. *Id.*

54. *Id.*

55. 896 F.2d at 1015.

56. *Id.* at 1016.

Defendant Epperly pleaded the Statute of Frauds as a defense to enforcement of the oral agreement.[57] The court, addressing this defense, found that "under Tennessee law, a wrongdoer is not permitted to rely on the Statute of Frauds defense."[58] The court further stated "the Statute of Frauds was 'enacted for the purpose of preventing fraud, and shall not be made the instrument of shielding, protecting, or aiding the party who relies upon it in the perpetration of a fraud . . . .'"[59] An even more basic notion regarding the Statute of Frauds, the court noted, was "its inapplicability to plaintiff's tort (i.e., fraud) claims."[60] Relying on *Haynes v. Cumberland Builders, Inc.*,[61] a decision of the Tennessee Court of Appeals, Middle Section, the court found the Statute of Frauds did not preclude a claim for fraud because "the Statute of Frauds applied only to contract actions and not to tort claims."[62]

The reasoning of the *Jarrett* court seems only reasonable in an age where the common business transaction is sealed with nothing more than an oral promise and a handshake.[63] To hold otherwise would be to give license to the perpetration of fraud and to allow the promisor who had no intent of honoring his promise to hide behind the Statute of Frauds.[64] Moreover, the *Jarrett* court's analysis reflects a proper application of the Statute of Frauds, in as much as its purpose is to prevent, rather than to perpetuate, fraud.[65]

Opponents of this reasoning, however, may argue that preventing the application of the Statute of Frauds to a tort action merely encourages the plaintiff-promisee to label his claim a tort, and thereby circumvent the Statute of Frauds.[66] This is not a terribly troubling concern, however, because merely labeling a cause of action a tort does not establish a valid cause of action for fraud.[67] The rigorously

---

57. *Id.* at 1017. Defendant contended the Statute of Frauds barred enforcement of the oral agreement because real estate was "embedded" in the agreement and the promises were incapable of being performed within one year. *Id.*

58. *Id.* at 1018.

59. *Id.* (quoting Interstate Co. v. Bry-Block Mercantile Co., 30 F.2d 172, 175 (W.D. Tenn. 1928)). *Accord* Cobble v. Langford, 230 S.W.2d 194 (Tenn. 1950).

60. Jarrett v. Epperly, 896 F.2d at 1019.

61. 546 S.W.2d 228 (Tenn. Ct. App. 1976).

62. *Jarrett*, 896 F.2d at 1019 (citing Haynes v. Cumberland Builders, Inc., 546 S.W.2d 228 (Tenn. Ct. App. 1976); *cf.* Webb v. Schultz, 198 S.W.2d 333 (Tenn. 1946) (in an action on the contract, "a false promise to sign an instrument in the future is not such fraud as would take the case out of the operation of the statute of frauds").

63. Stepaniuk, *supra* note 43, at 1245.

64. *See supra* notes 42-44 and accompanying text.

65. *See* Boutwell v. Lewis Bros. Lumber Co., 182 S.W.2d 1 (Tenn. Ct. App. 1944); Yates v. Skaggs, 213 S.W.2d 41 (Tenn. 1948).

66. Stepaniuk, *supra* note 43.

67. *Id.*

applied elements of a fraud action[68] ensure that the Statute of Frauds and its policy of preventing fraudulent enforcement of oral promises will not be easily circumvented.[69] An even greater guarantee of protection from circumvention of the Statute of Frauds exists with a promissory fraud claim than with a traditional fraud claim because the former requires proof that when the promisor made his promise he had no present intention of performing it.[70] As has been recognized since the days of Lord Bowen's famous pronouncement, proving another's present state of mind is no easy task.[71] In fact, although there is some question in Tennessee as to what burden of proof must be satisfied in an action for fraud,[72] Tennessee courts impose a greater burden of proof on plaintiffs alleging promissory fraud than traditional fraud.[73] Further, because fraud in Tennessee may never be presumed,[74] this burden on the party alleging promissory fraud

---

68. *See supra* note 13.

69. Stepaniuk, *supra* note 43.

70. *See* Fowler v. Happy Goodman Family, 575 S.W.2d 496 (Tenn. 1978).

71. *See supra* note 29 and accompanying text.

72. *See generally* Note, *Promissory Fraud in Tennessee*, *supra* note 38.

73. Sanders v. First National Bank, 114 Bankr. 507 (M.D. Tenn. 1990). In *Sanders*, Judge Wiseman quoted from Judge Conner's concurring opinion in Farmers & Merchants Bank v. Petty, 664 S.W.2d 77 (Tenn. Ct. App. 1983), where Judge Conner, regarding the burden of proof necessary to establish an action for promissory fraud, stated:

> I must agree with the majority's finding that the [s]upreme [c]ourt has indicated an unwillingness to apply the doctrine except in those cases where there is direct proof of a misrepresentation of actual *present* intention. There seems to be considerable reluctance on the part of the court to infer a false intent from the subsequent failure to follow through on a promise
>
> . . . .

In *Brungard* [where the (sic) found that plaintiff's recovery could be predicated on promissory fraud] there was overwhelming evidence that the defendant knew its statements regarding future intent were false at the time they were made. 608 S.W.2d at 588-90.

Sanders, 114 Bankr. at 516 (*quoting Farmers & Merchants Bank*, 664 S.W.2d at 82).

74. *See* Piccadilly Square v. Intercontinental Constr. Co., 782 S.W.2d 178 (Tenn. Ct. App. 1989); Groves v. Witherspoon, 379 F. Supp. 52 (E.D. Tenn. 1974); Snapp v. Moore, 2 Tenn. 236 (1814). Although fraud may never be presumed but must be proven, it may

> be inferred from the circumstances, such as the defendant's insolvency or other reason to know that he cannot pay, or his repudiation of the promise soon after it is made, with no intervening change in the situation, or his failure even to attempt any performance, or his continued assurances after it is clear that he will not do so.

PROSSER & KEETON, *supra* note 12, at 764-65. *See also* Edwards v. Travelers Ins., 563 F.2d 105, 112 (6th Cir. 1977) (applying Tennessee law) (Tennessee courts allow proof of fraud by wholly circumstantial evidence); RESTATEMENT (SECOND) OF TORTS § 530, comment d (1977): "The intention of the promisor not to perform an enforceable or unenforceable agreement cannot be established solely by proof of its nonperformance . . . ."

provides substantial protection against both the circumvention of the Statute of Frauds and spurious claims of promissory fraud.[75]

B. The Parol Evidence Rule

As with the applicability of the Statute of Frauds, questions as to the applicability of the Parol Evidence Rule arise where the fraud alleged is an oral promise made without the intent to perform it.[76] Under the Parol Evidence Rule, no proof of an oral agreement may be offered to vary the plain terms of a written contract.[77] Parol evidence, however, may be admissible to show that the agreement was entered into as a result of fraudulent inducement,[78] and in some cases to prove the actual intent of the parties when entering into an agreement.[79] The issue of admissibility of parol evidence, therefore, "arises when the opponent of the evidence of promissory fraud attempts to invoke the bar of the parol evidence rule."[80]

Again, as with the Statute of Frauds, the determination of the applicability of the Parol Evidence Rule to claims of promissory fraud involves a weighing of competing policy considerations: should the Parol Evidence Rule be applied to bar evidence of the oral promise in order to promote security and certainty in transactions, or should a plaintiff be allowed to introduce evidence of alleged oral promises with the understanding that merely introducing the evidence does not validly substantiate a claim for promissory fraud? It is generally agreed that the evidence is admissible because the Parol Evidence Rule does not bar admission of evidence of fraud.[81]

The explanation generally given for allowing parol evidence to prove fraud is that unless there is a *validly* formed contract embodied in a writing that the parties agree to be the final repository of their entire agreement, the integration aspect of the Parol Evidence Rule is inapplicable.[82] If the contract was not validly formed, i.e., fraud existed, then no contract exists to be protected by the exclusion of

---

75. *See* Stepaniuk, *supra* note 43.

76. *See* Note, *Promissory Fraud in Tennessee*, *supra* note 38.

77. Sanders v. First Nat'l Bank, 114 Bankr. 507, 517 (M.D. Tenn. 1990); Farmers & Merchants Bank v. Petty, 664 S.W.2d 77, 81 (Tenn. Ct. App. 1983); Dunn v. United Sierra Corp., 612 S.W.2d 470, 474 (Tenn. Ct. App. 1980).

78. Hill v. A.O. Smith Corp., 801 F.2d 217, 223 (6th Cir. 1986); Torbett v. Jones, 86 S.W.2d 898, 901 (Tenn. Ct. App. 1935).

79. Hill v. A.O. Smith Corp., 801 F.2d 217, 223 (6th Cir. 1986); *see* Little Darlin' Corp. v. Shelby Singleton Prod., Inc., 448 S.W.2d 447, 453 (Tenn. Ct. App. 1969).

80. Justin Sweet, *Promissory Fraud and the Parol Evidence Rule*, 49 CAL. L. REV. 877, 889 (1961) [hereinafter Sweet].

81. *Id.* at 877.

82. *Id.*

extrinsic evidence.[83] Because the existence of fraud constitutes a material defect in the formation of a contract, evidence of the fraud is admissible.[84] Further, as it is the general rule that a promise made without the intent to perform it is fraud,[85] evidence of such promissory fraud is generally admissible, despite the Parol Evidence Rule.[86]

Notwithstanding its tension with policies such as the Statute of Frauds and the Parol Evidence Rule, the doctrine of promissory fraud is a well recognized cause of action in modern American jurisprudence.[87] In fact, the Restatement (Second) of Torts recognizes promissory fraud as a valid cause of action:

> One who fraudulently makes a misrepresentation of fact, opinion, intention or law for the purpose of inducing another to act or to refrain from such action in reliance upon it, is subject to liability to the other in deceit for pecuniary loss caused to him by his justifiable reliance upon the misrepresentation.[88]

As Professor Keeton summarized, "All but a few courts regard misstatement of a present intention as a misrepresentation of a material fact; and a promise made without the present intent to perform it is held to be a sufficient basis for an action of deceit, or for restitution or other equitable relief."[89]

III. TENNESSEE'S POSITION REGARDING PROMISSORY FRAUD

Tennessee's stance regarding promissory fraud is unique. The Tennessee Supreme Court has been in a "holding" pattern since 1967 when the court, without formally adopting the majority position recognizing a cause of action for promissory fraud, announced its willingness to consider that position "in a proper case where justice demands."[90] Until then, Tennessee clearly subscribed to the minority view by refusing to recognize that promises made without the intent to perform them were actionable.[91] Despite the fact that the elements of promissory fraud have been designated[92] and applied by the Tennessee Court of Appeals,[93] the Tennessee Supreme Court has

---

83. *Id.*
84. *Id.*
85. *Id.* at 888. *See also supra* notes 21-23 and accompanying text.
86. Sweet, *supra* note 80, at 889.
87. James & Gray, *supra* note 15, at 507.
88. RESTATEMENT (SECOND) OF TORTS § 525 (1977).
89. PROSSER & KEETON, *supra* note 12, at 763.
90. Bolan v. Caballero, 417 S.W.2d 538, 541 (Tenn. 1967).
91. *See* A. Landreth Co. v. Schevenel, 52 S.W. 148 (Tenn. 1899).
92. *See* Brungard v. Caprice Records, Inc., 608 S.W.2d 585, 588 (Tenn. Ct. App. 1980).
93. *See id.*

never formally adopted the doctrine of promissory fraud.[94]

A. The Early Rule

Since it first addressed the issue of promissory fraud[95] in 1899, Tennessee has held the minority view that an action for fraud will not lie for representations or promises of future performance, even though the promise was made without the present intention to keep it.[96] In *A. Landreth Co. v. Schevenel*,[97] the Tennessee Supreme Court considered whether to set aside a settlement agreement between the parties.[98] The plaintiff, A. Landreth Company, alleged it was induced to enter the settlement agreement by Schevenel's repeated assurances and representations that it would continue its business as a grocery store and would resume its business relationship with the A. Landreth Company—neither of which Schevenel did.[99] The A. Landreth Company further alleged these assurances and representations were false and fraudulent, that they were known by Schevenel to be false and fraudulent, and that they were not performed as represented.[100] Finding the allegations of fraud "totally insufficient" to rescind the agreement, the court held "'[m]isrepresentations, in order to be fraudulent, must be of facts at the time or previously existing, and not mere promises for the future.'"[101] Stating the proposition that represents the early rule in Tennessee, the court further held "[a]n action for rescission for fraud cannot be predicated on a promise to do something in the future, although the party promising had no intention of fulfilling the promise at the time it was made."[102]

It has been suggested that the rationale underlying the *Schevenel* court's refusal to allow this action was fear of opening the door to superfluous litigation in fraud where there had been only a breach of contract.[103] Supporting this proposition is the *Schevenel* court's reliance on the principle set forth in *Farrar v. Bridges*,[104] wherein the Tennessee Supreme Court, addressing the insufficiency of proof of fraud, found:

---

94. Steed Realty v. Oveisi, 1991 WL 288197 at *3 (Tenn. Ct. App. May 8, 1991).

95. Although not labelled as promissory fraud the court considered an issue that mirrors the modern definition of promissory fraud.

96. *See* A. Landreth Co. v. Schevenel, 52 S.W. 148 (Tenn. 1899).

97. *Id.*

98. *Id.* at 148.

99. *Id.*

100. *Id.*

101. 52 S.W. at 148.

102. *Id.*

103. *See* Note, *Promissory Fraud in Tennessee*, *supra* note 38 at 324. *See supra* notes 30-35 and accompanying text.

104. 22 Tenn. 565 (1842).

> [N]o fraud is shown to distinguish this case from any other in which a party neglects or refuses to comply with his engagements. . . . Fraud, indeed, vitiates a contract into which it enters; but mere noncompliance with the terms of the contract, . . . is not fraud.[105]

If, in fact, this was the rationale underlying the *Schevenel* court's holding, its reasoning is skewed. It erroneously implies that allowing a fraud action as an alternative theory to a breach of contract action would impair the stability of contract law[106] by bringing about the merger of contract law and tort law into a single theory, leaving one indistinguishable from the other. This reasoning is unfounded.[107] As discussed earlier,[108] although the same set of facts may give rise to both an action on the contract and an action for fraud, these causes of action are not identical.[109] A plaintiff, therefore, should be allowed to plead them alternatively.[110] Moreover, none of the majority of states that recognize an action for promissory fraud has articulated that doing so impairs the stability of contract law.[111] Clearly, then, the *Schevenel* court erred in refusing to recognize promissory fraud on the grounds that doing so would result in litigation in which the party injured by simple breach of contract comes into court claiming fraud.

The *Schevenel* rule has been relied on in subsequent cases, confirming the early Tennessee position as one that denied a cause of action for promissory fraud.[112] In *Young v. Cooper*[113] the plaintiff, Young, sued to dissolve a partnership.[114] Young alleged he was induced to enter the partnership by fraudulent promises that the business would be expanded by ten stores within the next three years.[115] When one of the partners exercised his right to terminate the partnership before the expiration of the three-year term, Young sued, alleging the oral promise of expansion bound the partnership to a three-year commitment.[116] The Tennessee Court of Appeals,

---

105. Note, *Promissory Fraud in Tennessee*, *supra* note 38, at 324 (quoting *Schevenel*, 52 S.W. at 149).

106. *See* Polelle, *supra* note 14, at 575.

107. *See, e.g.*, Sunderhaus v. Perel & Lowenstein, 388 S.W.2d 140, 143 (Tenn. 1965) ("The alternative prayers of the bill for recision or for damages are not inconsistent.").

108. *See supra* notes 30-35 and accompanying text.

109. *See supra* notes 30-35 and 66-75 and accompanying text.

110. *See supra* note 107.

111. *See* Polelle, *supra* note 14, at 574-78.

112. *See* German-American Monogram Mfrs. v. Johnson, 182 S.W. 595 (Tenn. 1916); Pollock v. Bankson, 12 Tenn. App. 657 (1930) (in both cases defendant pled fraudulent inducement as an affirmative defense; in both cases the holding was based on grounds other than promissory fraud).

113. 203 S.W.2d 376 (Tenn. Ct. App. 1947).

114. *Id.* at 379.

115. *Id.* at 382.

116. *Id.* at 383.

Middle Section, held that the plans of expansion did not constitute a promise that the partnership would endure for the three-year period.[117] Relying on *Schevenel*, the court further found that if the plans did amount to promises of future events they were not actionable even though they proved to be false.[118]

B. The Recent Trend

The rule espoused in *Schevenel*[119] that promises of future events made with no present intent to perform them does not give rise to an action in fraud remained the rule in Tennessee on promissory fraud until 1967, when the Tennessee Supreme Court modified its position. In *Bolan v. Caballero*[120] the court, in dicta, announced its willingness to consider the majority rule on promissory fraud, stating, "[w]e recognize that the law in other jurisdictions is more generous towards plaintiffs in actions for fraud and deceit based on promissory misrepresentation than it is in Tennessee . . ., and in a proper case where justice demands we may be moved in that direction . . . ."[121]

In *Bolan*, Mrs. Bolan intended to sell her house to move to Florida and join her husband, who was stationed there.[122] Before her departure, Mrs. Bolan approached the defendants and proposed that they assume the mortgage on the house in exchange for which she would give up her equity in the property.[123] Apparently interested in the proposition, the defendants promised to determine what must be done to complete the transaction.[124] Mrs. Bolan departed for Florida, leaving defendants a forwarding address and expecting to hear from them.[125] A month later, having heard nothing from the defendants, Mrs. Bolan returned to Tennessee and approached the defendants in order to determine the status of the proposed transaction.[126] One of the defendants told Mrs. Bolan that she and her husband had been busy, and they had not prepared the necessary papers but would do so "as soon as possible."[127]

Several months later, Mrs. Bolan received word from the Federal Housing Administration that the property had been foreclosed on, bought by the defendants for considerably less than what was owed

---

117. *Id.*
118. *Id.*
119. 52 S.W. 148 (Tenn. 1899).
120. 417 S.W.2d 538 (Tenn. 1967).
121. *Id.* at 541 (citation omitted).
122. *Id.* at 539.
123. *Id.*
124. *Id.* at 539-40.
125. *Id.* at 540.
126. *Id.*
127. *Id.*

on the mortgage, and resold by the defendants at a substantial profit.[128] Mrs. Bolan then initiated an action against defendants in which she sought damages arising from the defendants' failure to keep their promise to enter into the proposed transaction.[129]

Affirming the chancellor, the Tennessee Supreme Court found that the defendants' statements did not constitute actionable fraud.[130] Moreover, the court, relying on the *Schevenel* rule, found that "[e]ven if the statement could be characterized as promissory fraud, which in our opinion it cannot be, it would be unenforceable . . . ."[131]

The court, however, announced its willingness to consider recognizing a cause of action for promissory fraud "in a proper case where justice demands."[132] The court reasoned this was not "a proper case" because Mrs. Bolan, the plaintiff, initiated the transaction, one of the defendants made no promise at all, and the other defendant's statements were not sufficiently fraudulent and inducing to be considered actionable.[133]

Although it did not adopt the doctrine of promissory fraud, the *Bolan* decision is significant because it represents the first step in Tennessee's movement toward recognizing promissory fraud as a valid cause of action. In *Fowler v. Happy Goodman Family*,[134] the movement gained momentum as the Tennessee Supreme Court, although finding insufficient evidence to establish promissory fraud, appeared willing to use the facts before it "as a vehicle for changing the rule with reference to 'promissory fraud' and to adopt the view followed by the majority of other jurisdictions."[135] In *Fowler* a group of musicians and their agent sued Fowler, their concert promoter, for fees they claimed were owed under contracts for musical performances.[136] In his answer, Fowler denied any breach of contract and raised the defenses of misrepresentation and fraud in the inducement but gave no statement of any facts supporting these defenses.[137] In response to plaintiffs' motion for a more definite statement as to the fraud and misrepresentation allegations, Fowler answered simply that "on information and belief" he had been induced to enter the contracts by false statements of intention and promises of future conduct by the plaintiffs.[138]

---

128. *Id.*
129. *Id.* at 538-39.
130. *Id.* at 540.
131. *Id.* at 541.
132. *Id. See supra* note 121 and accompanying text.
133. *Bolan*, 417 S.W.2d at 541.
134. 575 S.W.2d 496 (Tenn. 1978).
135. *Id.* at 499.
136. *Id.* at 497.
137. *Id.*
138. *Id.*

Affirming the decision of the lower court to grant plaintiffs' motion for summary judgment, the Tennessee Supreme Court found that the defendant's affidavit contained no evidence sufficient to sustain a defense of fraud in the inducement.[139] The court analyzed the promissory fraud issue by first recognizing the long-standing rule in *Schevenel* that in order to be actionable a fraudulent misrepresentation "must consist of a statement of an existing or past material fact, made with knowledge of its falsity or with reckless disregard of the truth."[140] The court also reaffirmed its statement in the *Bolan* decision where, eleven years earlier, the Tennessee Supreme Court noted its willingness "to consider adopting the rule followed in a majority of jurisdictions with respect to [promissory fraud] 'in a proper case where justice demands' . . . ."[141] The court further explained the majority position on promissory fraud, which states, "in order for actionable fraud to be based upon a promise of future conduct, it must be established that such a promise or misrepresentation was made with the intent not to perform. A statement of intention must be false and the intention not actually held."[142]

Although the *Fowler* court, like the *Bolan* court, expressed its willingness to adopt the majority position on promissory fraud where the facts demanded, the *Fowler* court nevertheless declined the opportunity to do so. The court found the facts legally insufficient to establish that the plaintiffs' promises and representations of their future conduct were made without the intent to perform them.[143] While the court found the plaintiffs' failure to perform the contracts was sufficiently stated, it found the defendant's subjective belief and his unspecified "information" legally insufficient to compel the adoption of the majority position on promissory fraud.[144]

The *Fowler* decision is an especially important case to the development of promissory fraud in Tennessee. In *Fowler* the Tennessee Supreme Court, although finding insufficient evidence to constitute promissory fraud, recognized the trend toward the majority rule.[145] More importantly, the court went on to judge the facts before it according to the majority rule.[146] Apparently, the only thing which prevented the court from expressly adopting the majority position on promissory fraud was the defendant's lack of proof.[147]

---

139. *Id.* at 499.
140. *Id.* at 498-99.
141. *Id.* at 499 (quoting *Bolan*, 417 S.W.2d at 541).
142. *Id.*, *citing* W. PROSSER, LAW OF TORTS § 109, at 728-30 (4th ed. 1971); RESTATEMENT (SECOND) OF TORTS § 530 (1977).
143. *Fowler*, 575 S.W.2d at 499.
144. *Id.*
145. *See supra* note 141 and accompanying text.
146. *See id.*; *supra* notes 142-44 and accompanying text.
147. *See supra* notes 143-44 and accompanying text.

Drawing closer to adopting the majority position on promissory fraud, the Tennessee Supreme Court denied certiorari[148] in a 1980 court of appeals decision in which the evidence supported plaintiff's allegation of promissory fraud. In *Brungard v. Caprice Records, Inc.*,[149] the Tennessee Court of Appeals, Middle Section, awarded plaintiff recovery on the alternative grounds of misrepresentation of a material existing fact and promissory fraud.[150] Despite language in the opinion indicating that the court's holding could have been based solely on traditional fraud grounds and that its analysis of promissory fraud was thus unnecessary,[151] some authority exists that *Brungard* represents Tennessee's adoption of the majority rule on promissory fraud.[152]

In *Brungard* an aspiring country music singer sued Caprice Records, alleging Caprice Records induced her to enter a recording contract by means of fraudulent misrepresentation.[153] Caprice Records produced custom records.[154] A custom record company sells only record production services—musicians, vocalists, studio time, tape, mixing mastering, and a producer.[155] Such a company typically does not provide promotional or marketing services sufficient to produce a commercially successful record or bear any financial risk if the record fails commercially.[156]

At trial the evidence demonstrated that Caprice Records induced the plaintiff to sign a custom record contract by falsely representing that it was taking a financial risk in signing her to a contract and that it would actively promote her record.[157] Specifically, Mr. Adams, a talent scout for Caprice Records, told the plaintiff that Caprice would invest $3,000.00 to her $2,966.00, when he knew Caprice would never invest any of its own money.[158] By this false statement,

---

148. *See supra* note 3 for a discussion of the effect of the supreme court's denial of certiorari.

149. 608 S.W.2d 585 (Tenn. Ct. App. 1980).

150. *Id.* at 590.

151. *See infra* notes 175-81 and accompanying text.

152. *See* Farmers & Merchants Bank v. Petty, 664 S.W.2d 77, 82 (Tenn. Ct. App. 1983) (Conner, J., concurring) ("First, I believe this court (and by denial of appeal, the supreme court) has already recognized the doctrine of promissory fraud in *Brungard v. Caprice Records* . . . ."); Holt v. American Progressive Life Ins., 731 S.W.2d 923, 927 (Tenn. Ct. App. 1987) ("our courts have now recognized a cause of action for promissory fraud") (citing Brungard v. Caprice Records, 608 S.W.2d 585 (Tenn. Ct. App. 1980). *But see* Farmers & Merchants Bank v. Petty, 664 S.W.2d 77, 80 (Tenn. Ct. App. 1983) ("It is to be noted that the [s]upreme [c]ourt has not adopted the doctrine of promissory fraud . . . .").

153. *Brungard*, 608 S.W.2d at 586.

154. *Id.* at 587.

155. *Id.*

156. *Id.*

157. *Id.* at 588.

158. *Id.* at 588-89.

Adams misled the plaintiff to believe that Caprice would make a financial commitment to promote her record which, in fact, it never did.[159] Moreover, Adams told the plaintiff Caprice was in the business of selling records when, in fact, only about $1,000.00 of its $761,000.00 gross income for that year was derived from selling records.[160] Adams also told the plaintiff Caprice's stock was publicly owned, leading her to believe Caprice was a large corporation and could therefore expend vast resources on the promotion and distribution of its records.[161] In fact, Caprice was owned by only two shareholders.[162] Adams further told the plaintiff Caprice had plans to build "Music World," a complex with a 3500 seat auditorium, to be used to expose new musical talent.[163] Caprice Records, however, actually had aborted plans to build "Music World" a year before Adams showed the plans to the plaintiff.[164] Finally, Adams told the plaintiff Caprice would release her record under its own well-known label, but Caprice instead released it under a label unknown in the industry.[165] Summarizing the transaction, the chancellor found

> [t]he entire presentation of Caprice to this plaintiff was designed to leave her with a false impression of the character of the transaction. In this production of a custom record, the plaintiff assumed all financial risks and paid all expenses. All she received was a record. Caprice intentionally misled her into believing that the transaction was something else—one in which the record company was taking a financial risk, had a financial interest in whether the record was a success, and one in which it would promote her record.[166]

The court of appeals began its analysis of the promissory fraud issue with a recognition of Tennessee's long-held rule in *Schevenel*[167] that an action for fraud cannot be based on representations or promises that something will be done in the future, even though the promise was made with no present intention to perform it.[168] Citing *Bolan*[169] and *Fowler*[170] the court then recognized the recent Tennessee trend toward the majority position recognizing a cause of action for promissory fraud.[171] Relying on these cases in which the Tennessee

---

159. *Id.* at 589.
160. *Id.*
161. *Id.*
162. *Id.*
163. *Id.*
164. *Id.*
165. *Id.*
166. *Id.* at 590.
167. *See supra* note 97 and accompanying text.
168. *Brungard*, 608 S.W.2d at 590.
169. *See supra* note 120 and accompanying text.
170. *See supra* note 134 and accompanying text.
171. *Brungard*, 608 S.W.2d at 590.

Supreme Court announced its willingness to consider adopting the majority position on promissory fraud where the party alleging the claim provides sufficient proof thereof, the court found that the evidence before it supported such a claim.[172] Thus, affirming the chancellor's rescission of the contract,[173] the court held "[p]laintiff's recovery therefore can be predicated on misrepresentation of a material existing fact and on misrepresentation of intention or promissory fraud."[174]

Notwithstanding the *Brungard* court's holding, the validity and necessity of its recognition of the doctrine of promissory fraud are questionable. Before its analysis of the promissory fraud issue, the court addressed plaintiff's claim of false misrepresentation in a commercial transaction.[175] Analyzing the misrepresentation made by Caprice Records to the plaintiff regarding its business activities and corporate structure, the court found that under Tennessee law these misrepresentations were material and that plaintiff relied on them in entering into the contract with Caprice Records.[176] "On this basis alone," the court stated, "plaintiff has proved her case."[177]

Having sufficient grounds upon which to affirm the chancellor, the court proceeded, arguably unnecessarily, with its analysis of promissory fraud.[178] Reaching the conclusion that promissory fraud was also applicable, the court held plaintiff could, in effect, have her choice of theories of recovery: traditional misrepresentation or promissory fraud.[179] Because, in the court's own words, recovery could have been based solely on misrepresentation of material existing fact,[180] the *Brungard* court apparently exceeded the scope of its review by analyzing and adopting the doctrine of promissory fraud. Further, because alternative theories of recovery underlie the *Brungard* deci-

---

172. *Id.*

173. The chancellor not only rescinded the contract but also awarded plaintiff treble damages and reasonable attorney's fees under the Tennessee Consumer Protection Act, TENN. CODE ANN. § 47-18-101 to -5002 (1988 & Supp. 1991). 608 S.W.2d at 587.

174. *Brungard*, 608 S.W.2d at 590.

175. *See id.* at 588. Under this theory, the court explained

> One, who in the course of his business, profession or employment, or in a transaction in which he has a pecuniary interest, supplies false information for the guidance of others in their business transactions, is subject to liability for pecuniary loss caused to them by their justifiable reliance upon such information if he fails to exercise reasonable care or competence in obtaining or communicating the information.

Id. at 588 (quoting Jasper Aviation, Inc. v. McCollum Aviation, 497 S.W.2d 240, 242 (Tenn. 1972) (emphasis omitted).

176. *See Brungard*, 608 S.W.2d at 590.

177. *Id.*

178. *See id.* at 590. *See supra* notes 167-74 and accompanying text.

179. *Brungard*, 608 S.W.2d at 590.

180. *Id.*

sion, *Brungard* cannot be said to represent Tennessee's unequivocal adoption of promissory fraud.[181]

C. *Actual Present Intent Not to Perform*

Bolan,[182] *Fowler*,[183] and *Brungard*[184] represent the recent Tennessee trend toward a willingness to adopt the majority position regarding promissory fraud "in a proper case where justice demands . . . ."[185] However, post-*Brungard* decisions reveal a troublesome aspect of promissory fraud that has kept subsequent Tennessee courts from wholly adopting the majority position: proof of the promisor's actual present intent not to perform his promise.

As discussed earlier, promissory fraud differs from traditional fraud in that the former requires proof that the promisor did not intend to honor his promise at the time he made it.[186] As was also discussed earlier and noted by Lord Bowen, "[i]t is true that it is very difficult to prove what the state of a man's mind at a particular time is, ..."[187] This truth has become especially evident in recent Tennessee decisions.

In *Farmers & Merchants Bank v. Petty*,[188] Petty executed notes in favor of plaintiff bank totalling more than $35,000.00.[189] Upon Petty's default, the bank sued for payment of the notes.[190] Petty then pled fraud and counterclaimed for damages, alleging he was induced to sign the notes by a promise from the bank's president that he would never have to repay the loans.[191] At trial the court dismissed plaintiff's suit and awarded Petty a judgment against the plaintiff for $5,000.00 and costs.[192] On appeal, the issue was whether material evidence existed sufficient to invalidate the note.[193]

The Tennessee Court of Appeals, Middle Section, initially noted that the Tennessee Supreme Court had not yet adopted the doctrine of promissory fraud but had "merely indicated a willingness to

---

181. *But see supra* note 152 and accompanying text.
182. *See supra* notes 120-33 and accompanying text.
183. *See supra* notes 134-47 and accompanying text.
184. *See supra* notes 148-81 and accompanying text.
185. Fowler v. Happy Goodman Family, 575 S.W.2d 496, 499 (Tenn. 1978); Bolan v. Caballero, 417 S.W.2d 538, 541 (Tenn. 1967); Brungard v. Caprice Records, Inc., 608 S.W.2d 585, 590 (Tenn. Ct. App. 1980).
186. *See supra* notes 33, 68-75 and accompanying text.
187. *See supra* note 71 and accompanying text.
188. 664 S.W.2d 77 (Tenn. Ct. App. 1983).
189. *Id.* at 78.
190. *Id.*
191. *Id.*
192. *Id.* at 79.
193. *Id.*

consider adopting the rule 'in a proper case where justice demands.'"[194] The court further stated that

> the [supreme] court has limited its willingness [to consider adopting promissory fraud] to cases where the statement of intention is shown to be false when made (i.e., a misrepresentation of actual present intention) by evidence other than subsequent failure to keep the promise or subjective surmise or impression of the promisee.[195]

Turning to the facts before it, the court found the bank's promise that Petty would never have to pay was not the equivalent of a promise that the bank would not sue Petty for non-payment.[196] Thus, reversing the judgment below, the court held "even if the 'guarantee' should be interpreted as an unequivocal promise that the bank would not insist upon payment . . . , there is no evidence, circumstantial or otherwise, that the representation was false when made, ..."[197] The court, therefore, found this fact situation was not one to which the supreme court had expressed a willingness to apply the doctrine of promissory fraud.[198]

The court in *Farmers & Merchants Bank* properly concluded evidence of the bank's present intent not to perform its promise was insufficient. The only evidence of the bank's intent Petty offered was a conversation between himself and the bank's president during which, Petty alleged, the bank's president "guaranteed me that I would never have it to pay."[199] Moreover, the court found evidence upon which to base the bank's president's alleged assurance. First, the note was secured by a second mortgage. As the court suggested, the bank's president reasonably may have told Petty he would not have to pay the note because the security was sufficient to cover it.[200] Second, Petty's son was the principal guarantor on the notes. As the court suggested, the bank may reasonably have expected that Petty's son would pay to protect him and Petty would thus not have to pay the notes.[201] The *Farmers & Merchants Bank* court, however,

---

194. *Id.* at 80. This finding by the court, and especially its use of the word "merely" in light of the earlier *Brungard* decision affirms this writer's contention, discussed *supra* at notes 174-80 and accompanying text, that *Brungard* cannot properly be said to represent Tennessee's adoption of the doctrine of promissory fraud.

195. Farmers & Merchants Bank v. Petty, 664 S.W.2d 77, 80-81 (Tenn. Ct. App. 1983).

196. *Id.* at 81.

197. *Id.*

198. *Id.*

199. *Id.* at 79. *See also* Stacks v. Saunders, 812 S.W.2d 587 (Tenn. Ct. App. 1990) (where the only proof plaintiff offered at trial was testimony of witnesses that defendant failed to fulfill his promise, evidence did not support a claim for promissory fraud).

200. *Farmers & Merchants Bank*, 664 S.W.2d at 81.

201. *Id.*

never suggested what degree of evidence would suffice to prove a promisor's actual present intent not to perform his promise. Even in *Brungard*,[202] where the court of appeals found ample evidence to support a claim of promissory fraud, the court did not set forth an objective standard or quantum of proof which must be satisfied to establish actual present intent not to perform. In this regard, Judge Conner explained that the "great reluctance" with which he concurred in *Farmers & Merchants Bank* concerned his belief that insufficient evidence existed of the present falsity of the bank's president's promise.[203] As he stated:

> I would be comfortable with a much more liberal approach to the doctrine of promissory fraud and the evidentiary requirements thereof. There are inherent and often insurmountable problems in proving "state of mind" at the time the representation is made as to its then truthfulness or falsity. Therefore, I would favor a construction that would allow an inference of original false intention to be made from evidence of subsequent actions.[204]

Applying a method of analysis very similar to Judge Conner's concurring analysis in *Farmers & Merchants Bank* (that is, inferring false intent based on evidence of subsequent actions),[205] the Tennessee Court of Appeals, Western Section, in *Maddux v. Cargill, Inc.*[206] concluded no evidence existed proving defendant's intent not to perform as stated.[207] In *Maddux*, Mr. Maddux enrolled in the payment-in-kind (PIK) program sponsored by the United States Department of Agriculture.[208] Under this program, Maddux agreed not to farm a portion of his land in exchange for a large quantity of high grade corn, which he was to pick up at a local distributor.[209] Upon arriving at the distributor to receive his corn, Maddux was told that before he received the corn he had to sign a certificate indicating he had received the quantity and quality of corn listed on the certificate.[210] Objecting to this procedure, Maddux allegedly signed the certificate only upon being assured he would receive the grade of corn he had been promised.[211] As it turned out, the defendant supplied Maddux with a far inferior grade of corn.[212] In his suit for fraudulent

---

202. 608 S.W.2d 585 (Tenn. Ct. App. 1980); *see supra* notes 148-74 and accompanying text.
203. *Farmers & Merchants Bank*, 664 S.W.2d at 82 (Conner, J., concurring).
204. *Id.* at 82-83. *See also supra* notes 72-74 and accompanying text.
205. *See supra* note 204 and accompanying text.
206. 777 S.W.2d 687 (Tenn. Ct. App. 1989).
207. *Id.* at 692.
208. *Id.* at 689.
209. *Id.*
210. *Id.*
211. *Id.*
212. *Id.* at 690.

misrepresentation, Maddux claimed the defendant made a promise to deliver a high grade of corn and subsequently delivered an inferior grade, which defendant represented to Maddux as being the promised high grade of corn.[213]

Holding the trial court improperly denied defendant's motion for directed verdict, the court found no evidence of a present intention on the part of the defendant to fail to perform as promised.[214] To bolster its conclusion, the court relied on evidence of subsequent events.[215] The court found that Maddux could have easily inspected the corn and on one occasion was offered the opportunity to dump the load of lower grade corn and receive another load.[216] Further, the evidence showed that Maddux knew he was receiving an inferior grade of corn and had even negotiated with the defendant to increase the quantity of corn in order to offset the quality deficiency.[217] Based on these facts, the court found no evidence of actionable fraud.[218]

As indicated, *Maddux* represents an effort by the Tennessee courts to apply a reasonable method of ascertaining the promisor's intent at the time he made his promise. Recognizing that fraud is difficult to prove, Tennessee courts allow proof thereof by "wholly circumstantial evidence."[219] Moreover, because proving one's present state of mind in an action for promissory fraud is even more difficult, the Tennessee courts, as evidenced by *Maddux*, will look to evidence of subsequent events as proof of the promisor's present intent.[220] Thus, the Tennessee courts have established a method for overcoming the most troublesome element of promissory fraud—proof of the promisor's present intent not to perform his promise.

Having established a method to deal reasonably with the most difficult element of promissory fraud, the Tennessee courts appeared poised to adopt the doctrine of promissory fraud, awaiting only the correct fact scenario. Supporting this conclusion, the district court from the Middle District of Tennessee announced in 1983, after a review of recent Tennessee case law addressing the issue of promissory fraud, that it was "convinced that the Tennessee Supreme Court would recognize a cause of action for promissory fraud under the

---

213. *Id.* at 692.
214. *Id.* at 693.
215. *Id.* at 692.
216. *Id.*
217. *Id.*
218. *Id.* at 693.
219. Edwards v. Travelers Ins., 563 F.2d 105, 112 (6th Cir. 1977) (applying Tennessee law).
220. *See supra* notes 215-18 and accompanying text. *See also* Sanders v. First Nat'l Bank, 114 Bankr. 507 (M.D. Tenn. 1990) (Because plaintiff could offer no evidence other than his own subjective beliefs about bank's intention, the court was unable to infer from bank's subsequent collection efforts that bank did not intend to honor security agreement.).

appropriate fact situation."[221] It is this writer's contention that such a fact situation presented itself in *Steed Realty v. Oveisi*,[222] and that *Steed Realty* therefore represents Tennessee's long-awaited and proper adoption of the doctrine of promissory fraud.[223]

IV. STEED REALTY

A. The Opinion

The plaintiff in *Steed Realty*, the seller of real estate (Steed), sued the purchasers of the land, alleging that they defaulted on their promissory notes for the purchase of the land.[224] The purchasers counterclaimed, alleging, *inter alia*, breach of contract, fraud, and misrepresentation.[225] The purchasers claimed Steed told them the area in which they bought land was a subdivision to be built in the future.[226] They claimed Steed subsequently promised that roads of ingress and egress to their property would be built and graveled according to county standards, and that electricity and water would be provided for their property.[227] The trial court found for the purchasers,[228] and Steed appealed, arguing, *inter alia*, that the trial court erred in finding promissory fraud was a viable theory of law in Tennessee.[229]

---

221. S&H Computer Sys., Inc. v. SAS Inst., Inc., 568 F. Supp. 416, 420 (M.D. Tenn. 1983).

222. 1991 WL 288197 (Tenn. Ct. App. May 8, 1991), *appeal denied*, Sept. 3, 1991.

223. *See supra* notes 5-7 and accompanying text. While some believe the Tennessee courts' adoption of promissory fraud occurred in Brungard v. Caprice Records, Inc., 608 S.W.2d 585 (Tenn. Ct. App. 1980), this writer thinks the *Brungard* court was too hasty in its recognition of promissory fraud. For authority that *Brungard* represents Tennessee's adoption of promissory fraud, see *supra* note 152 and accompanying text. For a discussion of this writer's disagreement with that position, see *supra* notes 175-81 and accompanying text.

224. *Steed Realty v. Oveisi*, 1991 WL 288197 (Tenn. Ct. App. May 8, 1991), *appeal denied*, Sept. 3, 1991.

225. *Id.* at *1 (The purchasers also alleged deceptive acts in violation of the Tennessee Consumer Protection Act.).

226. *Id.* at *3.

227. *Id.*

228. *Id.* At trial, the court dismissed Steed's suits against the purchasers to collect on the promissory notes. *Id.* The court also entered judgment against Steed for damages payable to each purchaser in the amount of injury each suffered. *Id.* The court also enjoined Steed from knowingly making any false promises regarding the sale of real estate for his business. *Id.* Moreover, the court held the contracts for the sale of the property and the accompanying promissory notes were void and the purchasers were to give Steed quit claim deeds to their property, pursuant to Steed paying the judgment. *Id.*

229. *Id.* at *4. Steed also appealed the issues of jurisdiction, the statute of limitations, the Tennessee Consumer Protection Act, and parol evidence. *Id.* at *3.

The court of appeals began its analysis of promissory fraud in a manner typical of prior Tennessee decisions. The court noted that although the Tennessee Supreme Court had never formally adopted the doctrine, the elements of promissory fraud had been set forth and applied by the Tennessee Court of Appeals.[230] The court then recognized the Tennessee Supreme Court's willingness to consider adopting the doctrine of promissory fraud "in the proper case where justice demands."[231] The court stated for an action of promissory fraud to succeed, the person alleging it must establish that the promise or representation was made with the intent not to perform it.[232] As an example of the application of this rule, the court cited the *Brungard* court's finding of sufficient evidence to support promissory fraud.[233] Reiterating the difficulty of proving promissory fraud, the court quoted from *Farmers & Merchants Bank*, where that court explained, "the standard for promissory fraud . . . is a stringent one and the court would have to re-write Tennessee law in order to make the facts of the case before it support the claim of promissory fraud."[234] After a review of the record from trial, the court affirmed the trial court's rescission of the contracts, holding "the doctrine of promissory fraud is appropriate in this case because Steed did not have the present intention to carry out certain promises at the time that he made them."[235]

B. The Rationale Behind Steed Realty

What motivated the *Steed Realty* court to adopt the doctrine of promissory fraud where other courts had previously declined the opportunity to do so? The court's rationale was unequivocal: Steed made promises to the purchasers that he had no present intent to perform at the time he made them.[236] As the court stated, "[t]he standard for proving promissory fraud . . . is that a representation must be made with the intent not to perform."[237] As discussed earlier, proof of the promisor's state of mind at the time he made his promise consistently has been the crucial element in an action for

---

230. *Id.* at *4 (citing Brungard v. Caprice Records, Inc., 608 S.W.2d 585 (Tenn. Ct. App. 1980)).

231. *Id.* (quoting Bolan v. Caballero, 417 S.W.2d 538, 541 (Tenn. 1967)).

232. *Id.* (citing Fowler v. Happy Goodman Family, 575 S.W.2d 396 (Tenn. 1978)).

233. *Id.* (citing Brungard v. Caprice Records, Inc., 608 S.W.2d 585 (Tenn. Ct. App. 1980)).

234. *Id.* at *4 (quoting Farmers & Merchants Bank v. Petty, 664 S.W.2d 77, 81 (Tenn. Ct. App. 1983)).

235. *Id.* at *6.

236. *Id.*

237. *Id.* at *5.

promissory fraud.[238] What, then, convinced the *Steed Realty* court that the purchasers satisfied this element? Relying on the record from trial, the court found Steed's testimony incredible in light of his previous convictions for criminal fraud, inconsistencies in his testimony, and his lack of candor.[239] As did the court in *Maddux*,[240] the *Steed Realty* court looked to evidence of Steed's subsequent actions to ascertain his state of mind at the time he made his promise. The trial court found his failure to make the improvements in the road or to provide electricity and water to the lots even after he repeated his promises to do so for several years as evidence of his intent not to perform.[241] The court of appeals also noted Steed's failure to offer any evidence explaining why he did not keep his promises, such as proof that he fell upon hard economic times or encountered some other subsequent condition rendering him unable to perform.[242] In fact, the evidence showed that Steed was still engaged in business and had a number of investments.[243] Moreover, as of the date of appeal, the promises were still unperformed, and no evidence existed indicating Steed intended to perform or was left unable to do so by circumstances beyond his control.[244]

The court of appeals found further evidence of Steed's false intent in the record. Steed testified he had no intention of providing electricity and water hook-ups because he was only obligated to provide the services mentioned in the sales contract.[245] Also, evidence showed that Steed referred to the property as a subdivision, implying that the area would become a residential area.[246] One purchaser testified Steed promised to build a clubhouse and turn it over to the landowners upon completion of the subdivision.[247] The court of appeals found that the purchasers were influenced to buy the lots by "the pretty picture Steed painted when he told them of the improvements to come."[248] Again, as evidence that Steed never intended to make these improvements, the court pointed to his testimony that

---

238. *See supra* notes 182-220 and accompanying text.

239. *Steed Realty*, 1991 WL 288197 at *5.

240. *See supra* notes 206-20 and accompanying text.

241. *Steed Realty*, 1991 WL 288197 at *5. The trial court also found that Steed's credibility was "substantially impeached" by his criminal record of fraudulent behavior, inconsistiencies in his testimony, and his lack of candor. *Id.*

242. *Id.*

243. *Id.* He had also made similar representations to other customers. *Id.*

244. *Id.*

245. *Id.* at *5. Steed also denied ever promissing to provide utility hook-ups. *Id.*

246. *Id* at *6. The court noted that most of the purchasers were buying for resort or retirement purposes, and the record shows that the lots could not be developed because of the lack of utilities and access to the property. *Id.*

247. *Id.*

248. *Id.*

"the contracts speak for themselves and [I am] not obligated outside the contracts."[249] Thus, convinced that the purchasers had met the standard for proving promissory fraud, i.e., had shown that Steed had no present intent to perform his promises, the court found the doctrine of promissory fraud applicable to this case.[250]

V. THE EFFECT OF *STEED REALTY*: CONCLUDING REMARKS

Before *Steed Realty* the Tennessee courts, although expressing a willingness to consider adopting promissory fraud, adhered to the minority position of refusing to recognize a cause of action for promissory fraud.[251] Various reasons existed for this position: confusion regarding the contract-tort distinction,[252] fear of contravening policies such as the Statute of Frauds[253] and the Parol Evidence Rule,[254] lack of an ascertainable standard of proof of promissory fraud,[255] and the absence of a "proper case where justice demand[ed]" adoption of the doctrine of promissory fraud.[256] Because it resolved many of these issues, *Steed Realty* represents Tennessee's proper adoption of the majority position on promissory fraud. As to the contract-tort distinction, the *Steed Realty* court properly recognized the purchasers' action as one for the tort of fraud rather than as one for breach of contract and thus recognized the two causes of action as separate and distinct.[257] As to the parol evidence issue, the *Steed Realty* court properly found that because the purchaser's recovery was based on fraudulent misrepresentation that induced the contract, rather than on a breach of contract theory, the Parol Evidence Rule did not apply.[258] As to the standard of proving promissory fraud, the *Steed Realty* court stated that the party alleging promissory fraud must prove "that a representation [was] made with the intent not to perform."[259] While the question still remains as to the degree of proof necessary, *Steed Realty* clearly presented "a proper case where justice demanded"[260] the court's

249. *Id.*
250. *Id.* The court followed the application of the doctrine of promissory fraud in Brungard v. Caprice Records, 608 S.W.2d 585 (Tenn. Ct. App. 1980). *Id*; *see supra* notes 149-81 and accompanying text.
251. *See supra* notes 90-94 and accompanying text.
252. *See supra* notes 30-35 and accompanying text.
253. *See supra* notes 38-75 and accompanying text.
254. *See supra* notes 76-86 and accompanying text.
255. *See supra* notes 182-204 and accompanying text.
256. Bolan v. Caballero, 417 S.W.2d 538, 541 (Tenn. 1967).
257. *Steed Realty*, 1991 WL 288197 at *7.
258. *Id.*
259. *Id.* at *5.
260. *See supra* note 257.

adoption of promissory fraud, as the evidence was overwhelming that Steed never intended to perform his promises.[261]

The court in *Steed Realty* based its holding solely on the grounds of promissory fraud. The court did not grant recovery on alternative grounds as did the court in *Brungard*, thereby raising uncertainty as to the court's position regarding promissory fraud.[262] Thus, it is this writer's contention that the Tennessee Supreme Court's denial of permission to appeal[263] in *Steed Realty* represents Tennessee's initial unequivocal adoption of the doctrine of promissory fraud.[264] Ending nearly a quarter of a century of waiting for "a proper case where justice demands" the adoption of promissory fraud,[265] *Steed Realty* represents Tennessee's recognition that promissory statements made with the present intent not to perform them are statements of present existing fact which, when relied upon, can result in legitimate and extensive injury to the promisee. As a result of the adoption of promissory fraud in *Steed Realty*, Tennessee courts now offer recourse to promisees injured by those "juggling fiends [who] palter . . . in a double sense."[266]

R. ALSTON HAMILTON

261. *See generally Steed Realty*, 1991 WL 288197 at *6.

262. *See* Brungard v. Caprice Records, Inc., 608 S.W.2d 585 (Tenn. Ct. App. 1980).

263. *See supra* note 3 and accompanying text.

264. *See supra* notes 180-81 and accompanying text.

265. It was in 1967, nearly twenty-five years ago, that the Tennessee Supreme Court in *Bolan* first expressed its willingness to adopt the doctrine of promissory fraud. *Bolan*, 417 S.W.2d at 541.

266. *See supra* note 1 and accompanying text.

THE FEDERAL CIRCUITS' RESPONSE TO CONFLICTING ARBITRATION AWARDS IN LABOR DISPUTES: SPLIT OR HARMONY BETWEEN THE SIXTH AND NINTH CIRCUITS?

> "[T]he arbitration promise is itself a contract. The parties are free to make that promise as broad or narrow as they wish, for there is no compulsion in the law requiring them to include any such promises in their agreement."[1]

INTRODUCTION

An employer contractually obligated to arbitrate individually grievances with two unions faces a dilemma when the source of a grievance concerns both unions.[2] Typically the dilemma appears when an employer assigns "new" work to employees in a particular bargaining unit, thereby eliminating participation of employees in another bargaining unit whose union then claims a right to the work under its collective bargaining agreement with the employer.[3] The excluded union then files a grievance that reaches arbitration.[4] If the arbitrator finds in favor of the complaining union, the other union may seek to force, through bipartite arbitration with the employer, a recognition that the original assignment of work was contractually mandated according to its agreement with the employer.[5] The possibility of conflicting arbitration awards can stain the sanctity the

---

1. United Steelworkers v. Warrior & Gulf Navigation Co., 363 U.S. 574 (1960) (Brennan, J., concurring).

2. This article is concerned with situations in which an employer's collective bargaining agreements with its unions require only bipartite arbitration. *See* Louisiana-Pacific Corp. v. International Bhd. of Elec. Workers, 600 F.2d 219 (9th Cir. 1979); Retail, Wholesale & Dept. Store Union, Local 390 v. Kroger Co., 927 F.2d 275 (6th Cir. 1991) (1991 *Kroger*); *see also Work Assignment Disputes Under the National Labor Relations Act*, 73 HARV. L. REV. 1150 (1960).

3. This is referred to as a "jurisdictional dispute," Transportation-Communication Employees Union v. Union Pacific R.R., 385 U.S. 157 (1966), or a "work assignment" dispute, Louisiana-Pacific v. International Bhd. of Elec. Workers, 600 F.2d 219 (1979); *see, e.g.*, Carey v. Westinghouse Elec. Corp., 375 U.S. 261, 263 (1964) (stating a jurisdictional dispute may include "a controversy as to whether certain work should be performed by workers in one bargaining unit or those in another.").

4. This assumes an arbitration clause exists in the collective bargaining agreement between employer and union.

5. Arbitration may be compelled by use of § 301 of the Labor Management Relations Act, 29 U.S.C. § 185 (1988).

law gives to private arbitration.[6] The role of tripartite arbitration[7] in this situation determines whether an employer will pay twice for the assigned work.[8]

The law in the Sixth Circuit allows a federal district court to order tripartite arbitration in the event of conflicting awards despite the absence of an express contractual duty of any party to participate.[9] The Ninth Circuit, by respecting the finality of arbitration decisions, allows conflicting awards to stand, even if it means an employer who has made conflicting promises must pay twice to have the work completed.[10]

The purpose of this article is to evaluate the reasoning of these decisions and to attempt to formulate a resolution of this apparent conflict. First, this will be accomplished by tracing the origins of the problem and second, by evaluating conflicting policy considerations to determine which should prevail. The result should enable employers and unions to predict a result, or better yet, to avoid the trap entirely.

DISCUSSION

I. The Policy Foundations

Arbitration of disputes over the application and meaning of labor contracts is unlike arbitration in other contexts. Section 301 of the Labor Management Relations Act[11] provides a means by which a

---

6. *See* Louisiana-Pacific Corp. v. International Bhd. of Elec. Workers, 600 F.2d 219 (1979) (forcing an employer to pay double the labor cost for new work assigned after conflicting arbitration awards granted from separate arbitrators).

7. Tripartite arbitration is also called "trilateral arbitration." In the context of this article, "tripartite arbitration" means a three-way arbitration between an employer and two unions. *See, e.g.*, Merton Bernstein, *Nudging and Shoving All Parties to a Jurisdictional Dispute Into Arbitration: The Dubious Procedure of National Steel*, 78 HARV. L. REV. 784 (1965) (discussing compulsion of arbitration).

8. *See* Louisiana-Pacific Corp. v. International Bhd. of Elec. Workers, 600 F.2d 219 (1979). For a spirited debate over the issue compare Bernstein, *supra* note 7 with Edgar A. Jones, *On Nudging and Shoving the National Steel Arbitration Into A Dubious Procedure*, 79 HARV. L. REV. 327 (1965) (discussing issues of consent to and compulsion of arbitration).

9. 1991 *Kroger*, 927 F.2d at 278; *see also* United Indus. Workers v. Kroger Co., 900 F.2d 944 (6th Cir. 1990) (1990 *Kroger*) (recognizing authority of a district court to order tripartite arbitration under certain circumstances).

10. *Louisiana-Pacific*, 600 F.2d at 223-24. A distinction must be drawn between ordering tripartite arbitration before and after the completion of bilateral arbitration. *See* National Post Office Mail Handlers v. American Postal Workers Union, 907 F.2d 190 (D.C. Cir. 1990).

11. 29 U.S.C. §§ 141-97 (1982).

party to a collective bargaining agreement may enforce its terms.[12] A distinct preference in favor of the existence of these contracts exists. In *United Steelworkers v. Warrior & Gulf Navigation Co.*,[13] the Supreme Court spoke to the importance of the agreement in the labor arena: "The collective bargaining agreement states the rights and duties of the parties. It is more than a contract; it is a generalized code to govern a myriad of cases which the draftsmen cannot wholly anticipate."[14]

In *Textile Workers Union v. Lincoln Mills*,[15] the Supreme Court called for the development of a body of federal common law in the arena of labor disputes. This growth was partially stunted when the Steelworkers Trilogy[16] removed the typical labor dispute from the judge and gave it to the arbitrator when an arbitration clause existed. Now, as long as an arbitrator grounds his opinion, even ambiguously, in the interpretation of the collective bargaining agreement, his interpretation and determination will be respected by the courts.[17] Crucial to an understanding of the dilemma is the recognition that parties to a collective bargaining agreement bind themselves voluntarily in the same way other private contracts are formed.[18] If an individual freely contracts into two conflicting arrangements, the law typically will not save him from himself.

The Supreme Court in *W.R. Grace & Co. v. Local Union 759, International Union of the United Rubber Workers*,[19] refused to save an employer from itself. The employer was faced with liability for

---

12. Section 301 of the Labor Management Relations Act states in relevant part:

> (a) Suits for violation of contracts between an employer and a labor organization representing employees in an industry affecting commerce as defined by this Act, or between such labor organizations, may be brought in any district court of the United States having jurisdiction of the parties, without respect to the amount in controversy or without regard to the citizenship of the parties.
>
> (b) . . . Any such labor organization may sue or be sued as an entity and in behalf of the employees whom it represents . . . .

29 U.S.C. § 185 (1982).

13. 363 U.S. 574 (1960). *Lincoln Mills* is part of the Steelworkers Trilogy, which also includes United Steelworkers v. American Mfg., 363 U.S. 564 (1960) and United Steelworkers v. Enterprise Wheel & Car Corp., 363 U.S. 593 (1960).

14. 363 U.S. at 578.

15. 353 U.S. 448 (1957).

16. *See supra* note 13.

17. United Steelworkers v. Enterprise Wheel & Car Corp., 363 U.S. 593 (1960) (holding an ambiguous arbitration determination is valid if arguably grounded in the collective bargaining agreement).

18. *See* United Steelworkers v. Warrior & Gulf Navigation Co., 363 U.S. 574, 580 (1960).

19. 461 U.S. 757 (1983).

violations of Title VII of the Civil Rights Act of 1964. In order to escape the liability, the employer entered into a conciliation agreement with the EEOC. The conciliation agreement required the employer to maintain a certain proportion of females in the event of a layoff. The requirement worked a result that conflicted with a seniority provision in the collective bargaining agreement with the union. The employer attempted to enjoin the subsequent arbitration of employee grievances that arose because of the conflict. The district court held the conciliation agreement should prevail. The Fifth Circuit reversed and forced the employer to arbitrate. The arbitrator granted back pay to the aggrieved employees. The Supreme Court held the arbitrator's award should be upheld despite its apparent conflict with the EEOC conciliation agreement. The Court stated: "The Company was cornered by its own actions, and it cannot argue now that liability under the collective-bargaining agreement violates public policy."[20]

The freedom of a union to agree or to refuse to agree to tripartite arbitration is conceptually inseparable from the freedom a union enjoys to bargain on all mandatory subjects.[21] Conceivably, an employer could negotiate for the inclusion of a tripartite arbitration clause in its collective bargaining agreements. Upon impasse, the union(s) could strike until the employer agreed to omit the tripartite arbitration clause from the collective bargaining agreement. After enduring the cost of a bitter strike a jurisdictional dispute could arise, possibly resulting in a pair of conflicting arbitration determinations. If a Sixth Circuit court is faced with this scenario, it can ignore the parties' ultimate bargain and force tripartite arbitration or stand a chance of running afoul of Sixth Circuit precedent by respecting the existing arbitration determinations despite the conflict.

The frequent focus of Section 301 litigation is the arbitration clause itself, which stems from the collective bargaining agreement.[22] As the Supreme Court acknowledged in *Carey v. Westinghouse Electric Corp.*,[23] "[T]he underlying objective of the national labor laws is to promote collective bargaining agreements and to help give substance to such agreements through the arbitration process."[24] The

---

20. *Id.*

21. Section 8(a)(5) of the Labor Management Relations Act prohibits an employer from refusing "to bargain collectively." Section 8(d) defines this duty as requiring an employer "to meet . . . and confer in good faith with respect to wages, hours, and other terms and conditions of employment." *See* NLRB v. Katz, 369 U.S. 736, 742 (1962).

22. *See* Nolde Bros. v. Local 358, Bakery & Confectionery Workers, 430 U.S. 243 (1977); Boys Markets v. Retail Clerks Union, Local 770, 398 U.S. 235 (1970); Textile Workers Union v. Lincoln Mills, 353 U.S. 448 (1957).

23. 375 U.S. 261 (1964).

24. *Id.* at 265 (quoting Carey v. Westinghouse Elec. Co., 184 N.E.2d 298 (N.Y. 1962) (Fuld, J., dissenting).

importance of the arbitration process in labor relations law rests with its ability to maintain industrial peace and reduce the risk of violence.[25]

Because of the need for alternative dispute resolution in the labor context, the federal law recognizes policy that strongly favors the finality of arbitration awards. As long as an arbitrator bases his or her award on the construction of the agreement itself, a court can enforce the award.[26] Hence, the role of the federal district court is limited to determining whether a claim made by an aggrieved party is, on its face, governed by the collective bargaining agreement.[27] In *United Steelworkers v. Enterprise Wheel & Car Co.*,[28] the Supreme Court upheld an ambiguous arbitration award on the basis that it reasonably could have been grounded in the agreement. This decision suggests the arbitrator's decision is given great weight. The Ninth Circuit's opinion in *Louisiana-Pacific v. International Brotherhood of Electrical Workers*,[29] in which the court upheld the integrity of conflicting arbitration awards and forced an employer to submit double payment for work, is rooted in the above policy.[30]

At odds with *Louisiana-Pacific* is the Sixth Circuit's opinion in *Retail, Wholesale & Department Store Union, Local 390 v. Kroger Co.*[31] In 1991 *Kroger* the Sixth Circuit rejected enforcement of a pair of conflicting bipartite arbitration awards in favor of ordering a new round of tripartite arbitration, thereby creating a stark conflict with the underpinnings of the *Louisiana-Pacific* decision.[32] The *Kroger* court simply stated that the "firm federal policy" of strictly honoring arbitration awards is not "served by blindly ordering enforcement of conflicting arbitration awards rendered in separate proceedings, neither of which had the ability to bind all interested parties."[33]

The *Louisiana-Pacific* and *Kroger* courts might not represent and espouse two distinct policies. Rather, as *Kroger* implies, the courts,

---

25. *See Lincoln Mills*, 353 U.S. at 445. The Court stated: "[Section 301] expresses a federal policy that our federal courts should enforce [arbitration] agreements on behalf of or against labor organizations and that industrial peace can be best obtained only in that way." *Id.*

26. *See* United Steelworkers v. Enterprise Wheel & Car Corp., 363 U.S. 593 (1960); Vic Wertz Distrib. Co. v. Teamsters Local 1038, 898 F.2d 1136 (6th Cir. 1990); Louisiana-Pacific Corp. v. International Bhd. of Elec. Workers, 600 F.2d 219 (9th Cir. 1979); *see also supra* note 17.

27. United Steelworkers v. Warrior & Gulf Navigation Co., 303 U.S. 574 (1960).

28. 363 U.S. 593 (1960).

29. 600 F.2d 219 (9th Cir. 1979).

30. *Id.* at 222. *Louisiana-Pacific* dealt with a district court's refusal to vacate two existing bipartite awards against an employer. The importance of giving the arbitrator's decision great deference is recognized only after arbitration.

31. 927 F.2d 275 (6th Cir. 1991).

32. *See id.* at 280.

33. *Id.*

following from the single broad-based policy of enforcing arbitration agreements, should carve a narrow exception when true impossibility of performance arises.[34] This approach shall be discussed later in this article.[35] Whether competing policies or an exception to the universal policy exists, this appears to be a distinction without a difference. If the latter applies, then the issue should be rephrased to ask whether an exception should exist.

II. Development of the Law

In *Carey v. Westinghouse Electric Corp.*,[36] the Supreme Court ordered to arbitration a dispute between the employer and the union that believed certain work should have been assigned to its members.[37] The *Carey* court ordered the bipartite arbitration between the union and employer. In his dissent, Justice Black spoke of the conflict: "Stripped of obscurantist arguments, this controversy is a plain, garden-variety jurisdictional dispute between two unions."[38] Acknowledging the quandary an employer faces, Justice Black continued: "[The employer] is trapped in a cross-fire between two unions. All he can do is guess as to which union's members he will be required . . . to assign disputed jobs. If he happens to guess wrong, he is liable to be mulcted in damages."[39]

In his concurrence in *Carey*, Justice Harlan spoke about the alternatives of granting jurisdiction with the Labor Board versus ordering arbitration. "[T]he choice in substance lies between a course which would altogether preclude any attempt at resolving disputes of this kind by arbitration, and one which at worst will expose those concerned to the hazard of duplicative proceedings."[40]

The *Carey* court never decided the propriety of a court-ordered tripartite arbitration. It did, however, provide both a majority opinion grounded in the policy underpinning *Louisiana-Pacific* and fertile ground for scholarly commentary concerned with avoiding the problem predicted by Justice Black.[41]

---

34. *See id.* The 1991 *Kroger* court relies on the nature of the arbitration award(s) as a key distinction from *Louisiana-Pacific*. *See id.* at 280-81.

35. *See infra*, notes 80-83 and accompanying text.

36. 375 U.S. 261 (1964).

37. *Id.* The dispute centered on whether "all production and maintenance" employees or "salaried technical" employees should have jurisdiction over the work. *See id.* at 262.

38. *Id.* at 274.

39. *Id.* at 275.

40. *Id.* at 273.

41. *See* Philip B. Kurland, *The Supreme Court, 1963 Term*, 78 HARV. L. REV. 143 (1964). In his article, Kurland states: "An employer can presumably protect himself from the ruling in *Carey* by contracting with each union for trilateral arbitration in the case of jurisdictional disputes giving rise to a demand for arbitration." *Id.* at 285; *see also* Bernstein, *supra* note 7; Jones, *supra* note 8.

Two years later the Supreme Court decided *Transportation-Communication Employees Union v. Union Pacific Railroad Co.*,[42] in which a majority, speaking through Justice Black, held that under the Railway Labor Act[43] the Railroad Adjustment Board must exercise its exclusive jurisdiction and include all affected parties when a job assignment dispute arises between an employer and its unions.[44] The *Union Pacific* court held a compulsory multipartite proceeding was necessary, even if one union declines to participate in any initial proceeding.[45]

The *Louisiana-Pacific* court's refusal to adhere to the holding in *Union Pacific* was grounded in the distinction that the *Louisiana-Pacific*'s arbitration proceeding found its source in contract and *Union Pacific*'s found its source in a statute.[46] The Supreme Court remanded *Union Pacific* to the Railway Adjustment Board, whose extensive jurisdiction over work assignment disputes flowed directly from an act of Congress rather than a private agreement.[47] The *Louisiana-Pacific* court found the "expansive purview" granted to the Railroad Adjustment Board to settle work assignment disputes was quite different from the authority of a private arbitrator to bind the parties that have not contracted with each other.[48] In *Kroger* the Sixth Circuit criticized the Ninth Circuit for its refusal in *Louisiana-Pacific* to recognize the importance of *Union Pacific* as valuable precedent when faced with jurisdictional work assignment cases.[49]

In *Columbia Broadcasting System v. American Recording and Broadcasting Ass'n*,[50] the Second Circuit held an employer may successfully maintain a Section 301 action under the Labor Management Relations Act[51] against two defendant unions to compel tripartite arbitration.[52] At the time of the action, there had been no arbitration determination or award made. The *Columbia Broadcasting* court recognized that while the language of Section 301 appeared to limit actions brought under it to suits "'for violation of contracts between an employer and a labor organization'"[53] the action could be maintained in light of decisions construing Section 301 very

---

42. 385 U.S. 157 (1966).
43. Railway Labor Act, 45 U.S.C. § 153 (1934).
44. *Union Pacific*, 385 U.S. at 157.
45. *Id.* at 165.
46. *Louisiana-Pacific*, 600 F.2d at 223-26.
47. *Union Pacific*, 385 U.S. at 164-65. The Court relied on 45 U.S.C. § 153.
48. *See Louisiana-Pacific*, 600 F.2d at 224.
49. 1991 *Kroger*, 927 F.2d at 280-81. Regarding its reasoning, the Sixth Circuit stated, "One of the reasons we hesitate to follow *Louisiana-Pacific* has to do with its reasons for rejecting . . . [Union Pacific]." *Id.* at 280.
50. 414 F.2d 1326 (2d Cir. 1969).
51. 29 U.S.C. § 185 (1982).
52. *Columbia Broadcasting*, 414 F.2d at 1329.
53. *Id.* at 1328.

broadly.[54] The *Columbia Broadcasting* court, however, purposely stopped short of determining whether the order was a proper exercise of the district court's power.[55]

In *Local 850, International Ass'n of Machinists v. T.I.M.E.-DC, Inc.*,[56] the Tenth Circuit decided a case that the 1991 *Kroger* court recognized as factually indistinguishable from those of its own.[57] *T.I.M.E.-DC* involved two separate bipartite arbitration orders against a single employer.[58] Each award required the employer to assign work to a different group of employees, rendering impossible a coextensive accommodation by the employer.[59] Refusing to accept the argument that the first union to arbitrate should prevail, the *T.I.M.E.-DC* court affirmed the district court's refusal to enforce a previous arbitration award as well as its referral of the dispute to tripartite arbitration.[60]

In 1990, the Sixth Circuit established the groundwork that eventually led to the 1991 *Kroger* opinion. In *United Industrial Workers v. Kroger Co.*,[61] the court held court-ordered tripartite arbitration was not appropriate before the completion of at least one bipartite arbitration proceeding.[62] The case involved a transfer of work from the first union to the second union.[63] The first union filed a grievance, and the employer issued an unsuccessful invitation to the second union to participate in a trilateral arbitration.[64] After the second union refused to participate, the employer became concerned that a pair of conflicting awards would result and therefore refused to arbitrate.[65] The first union filed a suit to compel arbitration, and the

---

54. *Id. See* John Wiley & Sons v. Livingston, 376 U.S. 543 (1964) (upholding a lower court's decision to compel arbitration despite the absence of a contract between the employer and union).

55. *Columbia Broadcasting*, 414 F.2d at 1329. "[T]hough the district court had jurisdiction of the case and had power to order a joint arbitration, the issue of whether there was a proper exercise of that power remains to be resolved." *Id.*

56. 705 F.2d 1275 (10th Cir. 1983).

57. *See* 1991 *Kroger*, 927 F.2d at 281.

58. *T.I.M.E.-DC*, 705 F.2d 1275. The dispute arose between the employer and two unions, the Machinists and Teamsters, over the assignment of yard "hostling and hookup." *See id.* at 1275-76.

59. *Id.* at 1276. The Sixth Circuit in the 1991 *Kroger* decision found the existence of this impossibility crucial in departing from *Louisiana-Pacific*. *See* 1991 *Kroger*, 927 F.2d at 280-81.

60. *T.I.M.E.-D.C.*, 705 F.2d at 1278.

61. 900 F.2d 944 (6th Cir. 1990) (1990 *Kroger*).

62. *Id.* at 948.

63. *Id.* at 945. The employer transferred work from the United Industrial Workers to the United Food and Commercial Workers (UFCW), therefore this is not a case in which "new" work was allocated. *Id.*

64. *Id.*

65. 1990 *Kroger*, 900 F.2d at 946.

employer interpled the second union in an attempt to obtain a tripartite arbitration order.[66]

After recognizing a duty to arbitrate on the part of the employer, the court stated "[a] possibility of external adverse consequences, such as exposure to inconsistent liabilities, cannot abrogate that duty."[67] The court continued by addressing whether tripartite arbitration could supplant that duty. Relying on *United States Postal Service v. American Postal Workers Union*,[68] the court held the required contractual nexus between the employer and the second union was lacking.[69] It may be, therefore, that until a conflict in awards exists, the court will not compel tripartite arbitration. The 1991 *Kroger* decision acknowledges the previous year's decision as establishing the right of the district court to order tripartite arbitration when a contractual nexus exists.[70]

III. The Inconsistency Between the Sixth and Ninth Circuits

A. Generally

In *Louisiana-Pacific,* the Ninth Circuit, speaking through a designated senior district court judge,[71] held an employer who failed to act to avoid conflicting bipartite arbitration awards essentially bore the risk of paying for both conflicting awards.[72] While the *Louisiana-Pacific* court acknowledged the "trend" favoring tripartite arbitration, it made clear that the principles underpinning the preference should not apply if two otherwise valid bipartite awards exist.[73]

Important in the reasoning of *Louisiana-Pacific* was the distinction between ordering tripartite arbitration before the award of conflicting arbitration and ordering it after the two awards had been made. According to the Ninth Circuit, if two awards have been handed down, their vacation is inappropriate.[74] Presumably, if one

---

66. *Id.* The employer relied on *Columbia Broadcasting* as authority to require tripartite arbitration. *Id.*

67. *Id.* at 946.

68. 893 F.2d 1117 (9th Cir. 1990).

69. 900 F.2d at 947.

70. 927 F.2d at 279. "Thus, we recognized the power to order tripartite arbitration but found the 'contractual nexus' lacking because Kroger's mere request to the UFCW to submit to tripartite arbitration did not constitute a grievance." *Id.*

71. The opinion was written by the Honorable M. Joseph Blumenfeld, Senior United States District Judge for the District of Connecticut, who sat by designation. 600 F.2d at 220.

72. *Id.* The court made it clear the onus is on the employer to "anticipate and to take steps to avoid" the problem. *Id.* at 224.

73. *Id.*

74. This appears to be the effect of the *Louisiana-Pacific* decision. *See* 600 F.2d at 219.

award is granted and another union files a grievance with the employer pursuant to the collective bargaining agreement but a second arbitration award is not made, a district court enjoys the power to order tripartite arbitration.

The 1991 *Kroger* court read the Ninth Circuit's refusal to follow *Union Pacific* in *Louisiana-Pacific* as a recognition that a party should not be compelled to engage in tripartite arbitration unless it expressly assumes the duty in its contract.[75] The 1991 *Kroger* court's interpretation led it to point out that *Louisiana-Pacific* did not remain true to its distinction because the Ninth Circuit intoned that the employer could have sought to force tripartite arbitration any time before the conclusion of both arbitrations.[76]

The crucial connection necessary to validate *Kroger*'s criticism lies in the assumption that *Louisiana-Pacific* really means what the 1991 *Kroger* court said it means. The *Louisiana-Pacific* court stated that "no contention is made that either of the agreements involved in this case contain any provision relating to tripartite arbitration of work assignment disputes."[77] However, it is more apparent that *Louisiana-Pacific* relied heavily on an employer's duty to avoid the problem and the federal policy of enforcing arbitration at almost any expense. The Ninth Circuit relied on the Steelworkers Trilogy:[78]

> [T]he grievance machinery under a collective bargaining agreement is at the very heart of the system of industrial self-government. Arbitration is the means of solving the unforeseeable by molding a system of private law for all the problems which may arise . . . . The processing of disputes through the grievance machinery is actually a vehicle by which meaning and content are given to the collective bargaining agreement.[79]

Thus it appears *Louisiana-Pacific*'s decision relied more heavily on the federal policy of enforcement rather than relying on the lack of a contractual duty to engage in tripartite arbitration. The danger may be that a subsequent court will read 1991 *Kroger* as a complete abandonment of *Louisiana-Pacific,* which may not be necessary.

---

75. *See* Retail, Wholesale & Dept. Store Union v. Kroger Co., 927 F.2d 275 (6th Cir. 1991).

76. 927 F.2d at 280. The Sixth Circuit stated:

> [T]he Ninth Circuit did not remain true to the distinction that it drew because later, in the same opinion, the court recognized that the employer could have forced tripartite arbitration, even though the parties had not agreed to tripartite arbitration in their collective bargaining agreements, had the employer acted before the arbitrators rendered separate awards.

Id.

77. 600 F.2d at 224.

78. *See supra* note 13.

79. 600 F.2d at 222.

In 1991 *Kroger*, the Sixth Circuit found *Union Pacific*'s logic more persuasive. The plaintiff, Retail Union, argued the firm federal policy of enforcing arbitration awards should prevail because the union "won the race" to arbitration and the award found its essence in the collective bargaining agreement.[80] The 1991 *Kroger* court disagreed. "Plaintiff's position is an oversimplification of the case and ignores the serious problems presented to the district court by conflicting demands for enforcement of conflicting arbitration awards."[81] This statement encapsulates the 1991 *Kroger* court's disagreement with *Louisiana-Pacific.*

Practically, the plaintiff's argument means the originally aggrieved union is in the better position in a work assignment dispute. Assuming the aggrieved union's claim has merit, and an arbitrator acknowledges this, the aggrieved union will always win the race to arbitration. The union originally receiving the work will have no reason to arbitrate until another award is made.

B. Nature of the Remedy

Louisiana-Pacific enforced two awards. The first ordered the assignment of work and award of back pay to the members of the aggrieved union.[82] The second award ordered recognition that the union that handled the work in the first place did so appropriately.[83] The simple distinction between *Louisiana-Pacific* and 1991 *Kroger* is that each arbitrator ordered Kroger to assign work to one union to the exclusion of the other. Essentially, the employer in *Louisiana-Pacific* could have complied with both awards, even though it was costly, but the employer in 1991 *Kroger* could not have complied. This distinction was made by the 1991 *Kroger* court. "Another reason for our declining to follow *Louisiana-Pacific* lies in a difference in the remedy awarded and enforced in that case . . . . [I]n this case compliance with both awards would be impossible."[84]

It may be that if the Sixth Circuit is faced with conflicting awards with which an employer could comply, it will follow *Louisiana-Pacific*. It is not evident whether the distinction in the nature of the remedy is strong enough to mute the Sixth Circuit's reliance on the policy supporting tripartite arbitration. If the nature of the remedy

---

80. 927 F.2d at 277.

81. *Id.*

82. 600 F.2d at 220. The arbitrator decided "[t]he tear down and reassembly of the generator portion of the machine in question was work within the jurisdiction of the Union. The Company shall forthwith pay an amount equivalent to the number of hours that such work took." *Id.* at 220 n.2.

83. *Id.* The second union, IBEW, feared the first arbitration award would impair its right to receive similar work in the future. *Id.*

84. 927 F.2d at 280-81.

survives as the most important distinction, then no conflict in the circuits exists.

A reasonable middle ground may lie in applying 1991 *Kroger* as the rule and *Louisiana-Pacific* as a carefully carved exception. *Louisiana-Pacific* could be harmoniously applied only when compliance with apparently conflicting awards remains theoretically possible.[85]

If this harmony is struck, it would necessarily preclude any deference to principles of equity. Neither circuit has given equitable consideration to exposure of conflicting awards. This omission leaves equity subservient to the respect the law gives to private contract principles. If an employer contracts into a potentially costly arrangement, the law appears to leave the risk with the employer. This makes contracting with two or more unions sticky business.

IV. Avoiding the Trap

A. The Employer's Perspective

The cases from the circuits tell us the best way to avoid conflicting awards is to seek the intervention of the federal district court before the completion of both arbitration procedures. Even *Louisiana-Pacific* recognized the availability of tripartite arbitration before the completion of two bipartite arbitrations.[86] In the Sixth Circuit, an employer may be able to approach the district court any time after the completion of one arbitration proceeding, but before completion of both arbitrations.[87] This may allow a court to find the contractual nexus required in the 1990 *Kroger* decision.

Another way to avoid back payment is to announce the availability of new jobs far enough in advance to allow the arbitration machinery to exhaust itself prior to the actual allocation of the work. This option would preclude double payment awards, as the work would not yet be available. Actual damages could not be sustained to that point, and arbitrators could at worst order the allocation to two separate bargaining units, as occurred in *Kroger*.

For the far-sighted, another option exists. If an employer negotiates for the inclusion of a tripartite arbitration clause in each collective bargaining agreement, the problem may be avoided altogether. A potential problem exists if the employer with two or more unions with which to contract fails to include the tripartite arbitration

---

85. This was the case in *Louisiana-Pacific*.

86. 600 F.2d at 225-26. "In addition to a contractual solution to this problem, the Company had the opportunity under Section 301 (a) of the Labor Management Relations Act . . . to require both unions to participate in an initial tripartite arbitration proceeding by seeking an order to do so from a federal court." *Id.*

87. The *Kroger* court may have left this possibility available.

clause in every collective bargaining agreement. At some point, the employer may seek to include a union whose collective bargaining agreement omits such a clause. The union is then called upon to engage in a practice that it specifically avoided during negotiations.

B. Union's Perspective

If the 1991 *Kroger* tripartite arbitration order applies only in cases where an employer's compliance with conflicting orders is impossible,[88] then a *Louisiana-Pacific* type of order may exist for unions in the Sixth Circuit when monetary awards conflict with assignment awards. However, a court faced with 1991 *Kroger* as precedent may choose to overlook the possibility that *Louisiana-Pacific* may act as an exception in instances when employer compliance with both awards is not impossible.

If assigned work has not yet begun and arbitration commences, the union may be better advised to avoid the time delay of two bipartite arbitrations followed by a court-ordered tripartite arbitration and proceed directly to voluntary tripartite arbitration. However, if work is assigned and begun, *Louisiana-Pacific* may serve as an avenue to monetary damages. Again, the Sixth Circuit law may develop to completely ignore *Louisiana-Pacific*.

CONCLUSION

There may or may not be a conflict in the circuits on the issue of court-ordered tripartite arbitration. If the Sixth Circuit chooses to recognize an enduring policy of avoiding any apparent conflicting award situations, employers will not face double payment to the extent they do in the Ninth Circuit. If the law, however, develops to recognize the importance of tripartite arbitration only when employer compliance with the awards is impossible, an employer may be exposed to double payment. The answer will not be known until the Sixth Circuit decides a case in which the employer faces orders that, while not particularly equitable, are not impossible to fulfill.

SHERRARD L. HAYES, JR.

88. *See supra* note 57 and accompanying text.

A SURVEY OF THE DECEMBER 1991 AMENDMENTS TO THE FEDERAL RULES OF CIVIL PROCEDURE

INTRODUCTION

The United States Supreme Court is responsible for promulgating general rules of practice and procedure as well as rules of evidence for the federal district courts and courts of appeals.[1] To meet this responsibility, the Chief Justice of the United States is authorized to summon annually the Judicial Conference of the United States[2] to consider administrative problems and policy issues affecting the federal court system and to recommend to Congress legislation affecting

1. *See* Rules Enabling Act §§ 1-2, 28 U.S.C. § 2072 (1988).

2. *See* 28 U.S.C. § 331 (1988). The Judicial Conference of the United States consists of the Chief Justice of the United States, the Chief Justice of the United States Court of Appeals for the Federal Circuit, the Chief Judge of the Court of International Trade, the chief judges of the other twelve United States Courts of Appeals and twelve district judges chosen for a term of three years by the judges of each circuit who meet at an annual judicial conference of the circuit. *Id. See* Thomas E. Baker, *An Introduction to Federal Court Rulemaking Procedure,* 22 TEX. TECH. L. REV. 323, 328 (1991).

the federal judicial system.[3] The Judicial Conference is assisted in recommending rule changes to Congress by the Committee on Rules of Practice and Procedure (Standing Committee).[4] The Standing Committee, in turn, is assisted by various Advisory Committees, who continuously study the Federal Appellate, Bankruptcy, Civil, and Criminal Rules.[5]

The process for amending federal rules is rather circuitous.[6] The particular Advisory Committee considering proposed rule changes publishes the proposals in order to give the public an opportunity to comment.[7] The Advisory Committee then publishes revised drafts, and the Standing Committee reviews the proposals and makes changes.[8] The Judicial Conference reviews the proposed amendments and forwards them to the United States Supreme Court, which adopts, modifies, or rejects the proposals.[9] The Supreme Court then transmits the adopted or amended rules to Congress by May first of the year in which the rule is to become effective.[10] If Congress takes no adverse action, the amended rule automatically becomes effective on December first of the year of transmittal.[11] On December 1, 1991, sixteen amendments to the Federal Rules of Civil Procedure emerged from this rulemaking obstacle course and became procedural law.[12] The amendments change Rules 5, 15, 24, 34, 35, 41, 44, 45, 48, 50, 52, 53, 63, 72, and 77. The rule changes apply to all actions commenced after December 1, 1991, and to all actions that were then pending "insofar as just and practicable."[13] The amendments change some of the rules only slightly while changing others substantially. This comment will examine the individual rules affected, their original purpose, how they have been amended, and the purpose and possible effects of the amendments.

---

3. Baker, *supra* note 2, at 328-29.

4. *See* 28 U.S.C. § 2073(b) (1988).

5. Baker, *supra* note 2, at 329.

6. *See* Gene R. Shreve, *Eighteen Feet of Clay: Thoughts on Phantom Rule 4(m)*, 67 IND. L.J. 85, n.2 (1991).

7. *See* Jack B. Weinstein, *Reform of Federal Court Rulemaking Procedures*, 76 COLUM. L. REV. 905, 908 (1976).

8. *Id.*

9. *Id.*

10. *See* 28 U.S.C. § 2074(a) (1988); *see also Amendments of Federal Rules of Civil Procedure*, 111 S. Ct. 813, 814 (1991) [hereinafter *Amendments*] (letter from Chief Justice Rehnquist transmitting amendments to Federal Rules of Civil Procedure to Congress).

11. *See* 28 U.S.C. § 2074(a) (1988); *see also* Walko Corp. v. Burger Chef Sys., 554 F.2d 1165, 1168-69 n.29 (D.C. Cir. 1977) (failure of Congress to suspend a proposed rule gives it the force of a regulation pursuant to the Rules Enabling Act rather than a legislative enactment).

12. The Supreme Court sent the proposed amendments to Congress on April 30, 1991. *See Amendments*, 111 S. Ct. at 814. Congress did not make any changes to the amendments. *See Changes to Federal Practice Take Effect; Civil Procedure*, MASS. LAW. WKLY., Dec. 2, 1991.

13. *Amendments*, 111 S. Ct. at 813.

Rule 5. Service and Filing of Pleadings and Other Papers

Rule 5 dictates the manner in which parties who are not in default for any failure of appearance are to be served with all papers and pleadings subsequent to the original complaint.[14] Rule 5 has two basic objectives. First, it seeks to insure a full exchange of written communications among the litigants so that each party has a copy of all papers.[15] Second, it attempts to establish a system for the filing of papers that will create an orderly court record.[16]

Paragraphs (a), (b), and (c) of Rule 5 have not been changed. These paragraphs direct when pleadings and papers must be served, how service is accomplished, and what procedure is followed when there are multiple defendants.[17] Paragraphs (d) and (e) of Rule 5 have been amended as follows:

> (d) **Filing; Certificate of Service.** All papers after the complaint required to be served upon a party, together with a certificate of service, shall be filed with the court within a reasonable time after service, but the court may on motion of a party or on its own initiative order that depositions upon oral examination and interrogatories, requests for documents, requests for admissions, and answers and responses thereto not be filed unless on order of the court or for use in the proceeding.[18]
>
> (e) **Filing with The Court Defined.** The filing of papers with the court as required by these rules shall be made by filing them with the clerk of the court, except that the judge may permit the papers to be filed with the judge, in which event the judge shall note thereon the filing date and forthwith transmit them to the office of the clerk. Papers may be filed by facsimile transmission if permitted by rules of the district court, provided that the rules are authorized by and consistent with standards established by the Judicial Conference of the United States. The clerk shall not refuse to accept for filing any paper presented for that purpose solely because it is not presented in proper form as required by these rules or any local rules or practice.[19]

The new Rule 5(d) still requires that all pleadings and papers subsequent to the complaint be filed within a reasonable time after being served on the parties.[20] The amended rule, however, now

---

14. *See* 4A CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1141 (1987) [hereinafter WRIGHT & MILLER]. Rule 4 governs service of the summons and original complaint. *See* FED. R. CIV. P. 4.
15. *See* WRIGHT & MILLER, *supra* note 14, § 1141.
16. *Id.*
17. *See* FED. R. CIV. P. 5.
18. FED. R. CIV. P. 5(d).
19. FED. R. CIV. P. 5(e).
20. *See* FED. R. CIV. P. 5(d).

requires that a certificate of service be filed with the court along with the court's copy of the papers served.[21] The former rule only required that the court receive a copy of the papers served.[22] Rule 5(d)'s requirement that a certificate of service be filed with the court generally has been imposed by local rule.[23] The purpose of this requirement is to put on file information concerning the date and manner in which service of a document was effected.[24] This information could be valuable if issues arise concerning effectiveness of service.[25]

The new Rule 5(e) retains the former rule's provision that all papers are considered filed with the court when they are placed in the possession of the clerk of court.[26] The rule, however, has been amended in two ways. First, the rule now prohibits court clerks from refusing to accept for filing any document solely because it is not in conformity with the federal rules or local rules or practices.[27] This amendment discontinues the practice by local clerks under the former rule of refusing to accept documents for filing simply because of technical or form mistakes.[28] Such a practice exposes litigants to the "hazards" of time bars and is not the proper role for the office of the clerk.[29] Enforcing the Federal Rules of Civil Procedure and local rules is the role for judicial officers only.[30] Second, Rule 5(e) has been amended to allow filing of documents by facsimile transmission (fax) if allowed by the local rules of the district court, provided the local rules conform with the standards promulgated by the Judicial Conference.[31]

---

21. *See* FED. R. CIV. P. 5(d); *Amendments to Federal Rules are Approved by Supreme Court*, 59 U.S.L.W. 2695 (1991).

22. *See* FED. R. CIV. P. 5, 28 U.S.C. §§ 567-68 (1982).

23. FED. R. CIV. P. 5 advisory committee's note.

24. *Id.*

25. *Id.*

26. *See* WRIGHT & MILLER, *supra* note 14, § 1153; *see also* Moffitt v. United States, 430 F. Supp. 34, 36 (E.D. Tenn. 1976) (complaint delivered to clerk of court within statute of limitations was considered properly "filed" with the court even though the complaint was not stamped as filed until after the statute of limitations had expired).

27. *See* FED. R. CIV. P. 5(e).

28. *See* FED. R. CIV. P. 5(e) advisory committee's note. It appears that the amended rule abolishes Rule 4(e) of the Federal District Court for the Western District of Tennessee and Local Rule 8(a)(1)(d) of the Federal District Court for the Middle District of Tennessee. These rules direct clerks of court in their district to refuse to file a document if the submitting party does not supply the correct number of copies required by the local rule. *Id. See* W.D. TENN. R. 4(e); M.D. TENN. R. 8(a)(1)(d).

29. *See* FED. R. CIV. P. 5 advisory committee's note.

30. *Id.*

31. *See* FED. R. CIV. P. 5(e). The Federal District Court for the Eastern District of Tennessee does not allow filing by facsimile transmission except with permission of the court. E.D. TENN. R. 5.1.

Rule 15. Amended and Supplemental Pleadings

"The objective of Rule 15 . . . is to allow the liberal use of amendments to . . . pleadings to facilitate the proper presentation of a case and to promote adjudication through litigation on the merits."[32] Rule 15 also allows parties to amend an original complaint for clarification or correction purposes "without being barred by a statute of limitations."[33] Restrictions on amendments to pleadings under Rule 15 are "often imposed in deference . . . to important policy concerns underlying the need for limitations periods"[34] such as preventing plaintiffs from bringing suit "after memories have faded, witnesses have died or disappeared, and evidence has been lost."[35]

Although Rule 15(c) has been revised significantly, it retains its primary purpose. Like the former rule, the new rule is based on the premise that once suit has been commenced the parties are not protected by the statute of limitations against a later amendment of defenses or claims arising out of the same conduct, transaction, or occurrence as the initial claim.[36] Rule 15(c) employs the legal fiction of relation back.[37] An amendment meeting the conditions imposed by the rule "relates back to the date of the original pleading."[38] Consequently, if an amendment relates back, it is treated as if it were filed along with the original complaint even if it actually was filed after the expiration of the applicable limitations period.[39] Moreover, if the original complaint was filed before the statute ran and if under the relevant statute of limitations the filing of a complaint counts as the commencement of the action, then relation back protects the postlimitations amendment from the limitations defense.[40]

---

32. *See* Lawrence A. Epter, *An Un-*Fortune*-ate Decision: The Aftermath of the Supreme Court's Eradication of the Relation-Back Doctrine*, 17 FLA. ST. U. L. REV. 713, 718 (1990). *See generally* 6 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1471 (1990).

33. *See* Epter, *supra* note 32, at 718.

34. *Id.* "The primary purpose of the statute of limitations defense is to compel the filing of a suit within a reasonable time period so that a defendant will have a fair opportunity to prepare his defense" without prejudice. Statutes of limitation also further the policy of judicial economy. Without statutes of limitation, courts would bear the expense of ascertaining antiquated factual issues in the form of wasted court time and resources. *See* Nathan M. Gundy, III, Note, Schiavone v. Fortune: *A Clarification of the Relation Back Doctrine*, 36 CATH. U. L. REV. 499, 505-06 (1987).

35. *See* Chase Secs. Corp. v. Donaldson, 325 U.S. 304, 314 (1945).

36. *See* WRIGHT & MILLER, *supra* note 32, § 1496.

37. *See* FED. R. CIV. P. 15(c).

38. *Id.*

39. *See* Robert Brussack, *Outrageous* Fortune: *The Case for Amending Rule 15(c) Again*, 61 S. CAL. L. REV. 671, 674 (1988).

40. *Id.*

The former rule allowed relation back if a claim or defense asserted in the amended pleading arose out of the "conduct, transaction, or occurrence" set forth in the original pleading.[41] Also, under the former rule, if a complaint named the wrong party or omitted a party from a pleading, the error or omission could be corrected if (1) the claim asserted arose out of the "conduct, transaction, or occurrence" set forth in the original pleading and (2) the party had received notice of the institution of the action or should have known that, but for a mistake, the action would have been brought against that party.[42] Notice of the institution of the suit or of a mistake in pleading must have been received by the party "within the period provided by law for commencing the action against the party" for relation back to occur.[43]

The amended Rule 15(c) now reads:

> (c) **Relation Back of Amendments.** An amendment of a pleading relates back to the date of the original pleading when
>
> (1) relation back is permitted by the law that provides the statute of limitations applicable to the action, or
> (2) the claim or defense asserted in the amended pleading arose out of the conduct, transaction, or occurrence set forth or attempted to be set forth in the original pleading, or
> (3) the amendment changes the party or the naming of the party against whom a claim is asserted if the foregoing provision (2) is satisfied and, within the period provided by Rule 4(j) for service of the summons and complaint, the party to by brought in by amendment (A) has received such notice of the institution of the action that the party will not be prejudiced in maintaining a defense on the merits, and (B) knew or should have known that, but for a mistake concerning the identity of the proper party, the action would have been brought against the party.
>
> The delivery or mailing of process to the United States Attorney, or the United States Attorney's designee, or the Attorney General of the United States, or an agency or officer who would have been a proper defendant if named, satisfies the requirement of subparagraphs (A) and (B) this paragraph (3) with respect to the United States or any agency or officer thereof to be brought into the action as defendant.[44]

---

41. *See* FED. R. CIV. P. 15(c), 28 U.S.C. app. 549 (1982).
42. *Id.*
43. *Id.*
44. FED. R. CIV. P. 5(c).

The rule was amended to prevent parties against whom claims are made from taking unjust advantage of inconsequential errors in pleadings in order to mount a statute of limitations defense.[45] Additionally, paragraph (c)(1), a new provision, clearly states that Rule 15 does not preclude any relation back that is permitted under any state or federal limitations law.[46] Therefore, if state or federal law provides a more lenient principle of relation back than Rule 15, then that law is available to save the claim.[47] State law usually will supply the applicable statute of limitations.[48] When federal jurisdiction is based on diversity of citizenship, the state in which the district court sits supplies the specific limitations law, and that law also determines whether service must be effected within the statute of limitations.[49] Even where the district court's jurisdiction is based on a federal question, if the applicable federal law provides no guidance with regard to the expiration of the statute of limitations, state laws of limitation may govern.[50] While federal courts may borrow a state's statute of limitations, the federal court should borrow no more of the statute than necessary to fill the gap left by Congress.[51] The borrowed state statute of limitations' provisions for service may be ignored because Federal Rules of Civil Procedure 4(a) and 4(j) provide for service.[52] Regardless of whether state or federal law applies, Rule 15(c)(1) now provides that if the applicable law affords a more forgiving principle of relation back than Rule 15, it should be used.[53]

The changes in paragraph (c)(2) only alter the structural form of the relation back requirement as provided in the former Rule 15.

---

45. *See* FED. R. CIV. P. 15 advisory committee's note.

46. *Id.*

47. *Id. See, e.g.*, Marshall v. Mulrenin, 508 F.2d 39 (1st Cir. 1974) (state statute permitted the addition of parties as defendants to relate back for the purpose of the state's statute of limitations in spite of contrary provisions of Rule 15).

48. *See* FED. R. CIV. P. 15 advisory committee's note.

49. *See* Walker v. Armco Steel Corp., 446 U.S. 740, 742 (1980) (In the absence of a federal rule directly on point, state service requirements that are an integral part of the state statute of limitations should control in a federal suit based on diversity.).

50. *See* Board of Regents v. Tomanio, 446 U.S. 478, 484 (1980) (Where federal law provides no rule of decision for actions brought under federal law, the appropriate state law may be "borrowed" if it is not inconsistent with the federal policy underlying the cause of action at hand.); *see also* Robertson v. Wegmann, 436 U.S. 584, 588 (1978) (federal courts may only disregard state law inconsistent with federal law or the Constitution); Johnson v. Railway Express Agency, 421 U.S. 454, 465 (1975) (state statute of limitations govern except when inconsistent with the federal policies supporting the cause of action). In other situations involving federal questions, however, the controlling limitations law may be federal law. *See* FED. R. CIV. P. 15 advisory committee's note.

51. *See* West v. Conrail, 481 U.S. 35, 39-40 (1987).

52. *Id.*

53. FED. R. CIV. P. 15 advisory committee's note.

Rule 15(c)(2) retains the old Rule 15(c) allowance of relation back where the claim or defense asserted in the amendment arose out of the same claim, transaction, or occurrence alleged in the original pleading.[54]

Paragraph (c)(3) was amended to change the result in the United States Supreme Court case of *Schiavone v. Fortune*.[55] In *Schiavone*, three plaintiffs filed libel actions on May 9, 1983, in response to the publication of the cover story in *Fortune* magazine appearing May 31, 1982.[56] The applicable New Jersey statute of limitations required a libel action to be brought within one year of the publication of the alleged libel.[57] The complaints, which only listed "Fortune" as defendant, were filed within the limitations period.[58] *Fortune*, however, was only a trademark and the name of an internal division of Time, Incorporated.[59] When the appropriate agent of Time, Incorporated received the complaints, the agent refused service simply because Time was not named as defendant in the complaints.[60] On July 19, 1983, well after the statute of limitations had run, the plaintiffs amended their complaint to name "Fortune, also known as Time, Incorporated" as defendants.[61] Time moved to dismiss the complaints on the ground that it had become a defendant only after the statute of limitations had run.[62] The district court concluded that the amendments to the complaints did not relate back to the filing of the original complaint because Time did not receive notice of the institution of the suit within the limitations period.[63] The Third Circuit agreed.[64]

The Supreme Court affirmed in a divided opinion.[65] The majority[66] maintained that notice to Time did not occur "within the period provided by law for commencing the action against" Time as required by Rule 15 because Time received notice only after the expiration of the one-year statute of limitations.[67] The dissent[68] argued that Time

---

54. *See* FED. R. CIV. P. 15(c)(2).
55. 477 U.S. 21 (1986). *See* FED. R. CIV. P. 15 advisory committee's note.
56. *Schiavone*, 477 U.S. at 22.
57. *Id.* at 23. *See* N.J. STAT. ANN. § 2A:14-3 (West 1952).
58. *Schiavone*, 477 U.S. at 22-23.
59. *Id.* at 23.
60. *Id.*
61. *Schiavone*, 477 U.S. at 23.
62. *Id.*
63. *Id.* at 24.
64. *See* Schiavone v. Fortune, 750 F.2d 15 (3d Cir. 1984).
65. Schiavone v. Fortune, 477 U.S. 21, 32 (1986).
66. Writing for the majority, Justice Blackmun was joined by Justices Brennan, Marshall, Powell, Rehnquist, and O'Connor. *Id.* at 22.
67. Justice Blackmun determined that "'within the period provided by law for commencing the action'" meant "'within the applicable limitations period.'" *Id.* at 30-31 (quoting FED. R. EVID. 702 advisory committee's note).
68. Justice Stevens dissented and was joined by Chief Justice Burger and Justice White. *Id.* at 22.

received notice of the suit within the time allowed for service of process under Rule 4(j) and that there was no evidence Time was prejudiced by the amendment.[69] The majority held the language of Rule 15(c) rendered the plaintiffs' complaints untimely even though they were filed within the statute of limitations and even though Time, Incorporated clearly had adequate notice of the timely filed complaints.[70]

The revised Rule 15(c) changes the result in *Schiavone* by providing that if an intended defendant receives notice of the commencement of an action within the 120 days provided for service of process under Rule 4(j),[71] a complaint may be amended at any time to correct a formal defect such as misnomer or misidentification.[72] This relaxation of the relation back requirement effectively reverses the Supreme Court's "unduly restrictive reading"[73] of Rule 15(c). Moreover, this relaxation embraces the principle purpose of the rule by allowing the plaintiff to correct a pleading error after the statute of limitations has run if the correction will not prejudice the opposition in any way.[74]

The final paragraph to the new Rule 15(c) is designed to produce results contrary to those in cases where an action against a government official or agency was dismissed simply because the wrong government defendant was named in the original complaint.[75] The

---

69. *Id.* at 34.

70. *See Schiavone*, 477 U.S. at 30-31 (plain language indicates amendment must be made within limitations period).

71. In the original amended Rule 15 sent to Congress, Rule 15 referred not to Rule 4(j) but to a nonexistent Rule 4(m). *See* Gene R. Shreve, *Eighteen Feet of Clay: Thoughts on Phantom Rule 4(m)*, 67 IND. L.J. 85, 87 (1991). The Judicial Conference proposed revisions to Rule 4 and Rule 15 simultaneously. Proposed Rule 4 had moved the 120-day service limit to a proposed paragraph "m." When Congress did not submit proposed Rule 4 to the Supreme Court, the reference to paragraph "m" in the proposed Rule 15, unexplainably, was not corrected to "j." *Id.* The mistake has been corrected by congressional action. Act of December 9, 1991, Pub. L. No. 102-198, 105 Stat. 162 (1991).

72. *See* FED. R. CIV. P. 15 advisory committee's note.

73. *See* Joseph P. Bauer, Schiavone: *An Un*-Fortune-*ate Illustration of the Supreme Court's Role as Interpreter of the Federal Rules of Civil Procedure*, 63 NOTRE DAME L. REV. 720 (1988).

74. *See Schiavone*, 477 U.S. at 38 (Stevens, J., dissenting).

75. *See* Rys v. United States Postal Serv., 886 F.2d 443 (1st Cir. 1989) (plaintiff, a postal worker appearing pro se, not allowed to add Postmaster General as defendant after statute of limitations expired); Gardner v. Gartman, 880 F.2d 797 (4th Cir. 1989) (plaintiff, an employee of the Department of the Navy, not allowed to add Secretary of the Navy as defendant after statute of limitations expired); Williams v. Army and Air Force Exch. Serv., 830 F.2d 27 (3d Cir. 1987) (*Schiavone* does not permit relation back of motion to join Secretary of Defense and knowledge of suit may not be imputed from named defendant to Secretary); Bell v. Veterans Admin. Hosp., 826 F.2d 357 (5th Cir. 1987) (plaintiff named hospital as defendant in sexual harassment suit but court concluded Administrator

rule now allows relation back of motions to add any officer or agency of the United States government if service was effected on a United States Attorney, a United States Attorney's designee, the Attorney General of the United States, or an agency or officer who would have been a proper defendant if named.[76]

Rule 24. Intervention

Rule 24 allows a nonparty to intervene in a suit upon timely application.[77] Rule 24 has two important purposes: "to foster economy of judicial administration and to protect nonparties from having their interests adversely affected . . . without their participation."[78] Rule 24(a) grants a nonparty the ability to intervene in an action as a matter of right under certain conditions.[79] If a statute of the United States grants an unconditional right to intervene or if the applicant party claims an interest relating to the subject of the action and the party cannot protect his or her interest through the existing party, intervention must be granted.[80] In contrast, intervention through Rule 24(b) is discretionary with the court.[81] If a statute of the United States grants a conditional right to intervene or if the applicant's claim or defense has questions of fact or law in common with the main action, the court may allow intervention.[82]

Rule 24(c) has been amended and now provides:

> (c) A person desiring to intervene shall serve a motion to intervene upon the parties as provided in Rule 5. The motion shall state the grounds therefor and shall be accompanied by a pleading setting forth the claim or defense for which intervention is sought. The same procedure shall be followed when a statute of the United States gives a right to intervene. When the constitutionality of an act of Congress affecting the public interest is drawn in question in any action in which the United States or an officer, agency, or employee thereof is not a party, the court shall notify the Attorney General of the United States as provided in Title 28, U.S.C. § 2403. When the constitutionality of any statute of a State affecting the public interest is drawn in question in any action in which that State or any agency, officer, or employee thereof is not a party, the court shall notify the attorney general of the State as provided

---

of Veteran Affairs was proper party and refused to permit relation back of plaintiff's joinder).

76. *See* FED. R. CIV. P. 15 advisory committee's note.

77. *See* FED. R. CIV. P. 24. *See generally* 7 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1901 (1986).

78. Stallworth v. Monsanto Co., 558 F.2d 257, 265 (5th Cir. 1977).

79. *See* FED. R. CIV. P. 24(a); *see also* WRIGHT & MILLER, *supra* note 77, § 1906.

80. FED. R. CIV. P. 24(a).

81. *See* FED. R. CIV. P. 24(b); *see also* WRIGHT & MILLER, *supra* note 77, § 1913.

82. FED. R. CIV. P. 24(b).

> in Title 28, U.S.C. § 2403. A party challenging the constitutionality of legislation should call the attention of the court to its consequential duty, but failure to do so is not a waiver of any constitutional right otherwise timely asserted.[83]

All of the former Rule 24(c) has been retained by the amended rule. As under the former rule a person who wishes to intervene into a suit must serve on the existing parties a motion to intervene stating the grounds for intervention and the appropriate pleadings setting forth the particular claims or defenses.[84] The new rule also retains the former rule's provision that when a suit challenges the constitutionality of an act of Congress, the court must notify the Attorney General of the United States to allow the United States an opportunity to intervene.[85]

The amendment to Rule 24(c) consists of the two sentences added to the end of the former rule. The new language requires the court to notify the attorney general of the state in which the district court sits when a suit questions the constitutionality of a state statute affecting the public interest and the state government is not a party.[86] The amendment thus requires state governments to be informed when their statutes are called into question and facilitates the state's intervention into such suits for the defense of the statutes in question as provided by 28 U.S.C. section 2403.[87]

The amended rule places the burden of notification squarely on the court as required by 28 U.S.C. section 2403.[88] The purpose of this added language is to resolve any confusion created by local district court rules.[89] Although the court bears the burden of notifying the appropriate government official, the rule directs the party challenging the constitutionality of the statute to remind the court of this duty.[90] The rule protects the constitutional rights of parties who

---

83. FED. R. CIV. P. 24(c).
84. *See* FED. R. CIV. P. 24(c).
85. *Id.*
86. *See* FED. R. CIV. P. 24(c).
87. *See* 28 U.S.C. § 2403 (1988). Section 2403(b) provides:

> (b) In any action, suit, or proceeding in a court of the United States to which a State or any agency, officer, or employee thereof is not a party, wherein the constitutionality of any statute of that State affecting the public interest is drawn in question, the court shall certify such fact to the attorney general of the State, and shall permit the State to intervene for presentation of evidence, if evidence is otherwise admissible in the case, and for argument on the question of constitutionality. The State shall, subject to the applicable provisions of law, have all the rights of a party and be subject to all liabilities of a party as to court costs to the extent necessary for a proper presentation of the facts and law relating to the question of constitutionality.

88. *Id.*
89. *See* FED. R. CIV. P. 24 advisory committee's note.
90. *See* FED. R. CIV. P. 24(c).

are not aware of the notice requirement by providing that the failure of a party to call the court's attention to its duty cannot constitute a waiver.[91]

Rule 34. Production of Documents and Things and Entry Upon Land for Inspection and Other Purposes

"The purpose of Rule 34 is to make relevant and non-privileged documents and objects in the possession of one party available to the other . . . ."[92] This eliminates strategic surprise, helps simplify issues, and expedites the trial process.[93] In general, that which is discoverable under Rule 26[94] is discoverable under Rule 34.[95] Rule 34(a) allows a party to request another party to produce "documents" and "tangible things" and grants a party "entry upon designated land or other property" for discovery purposes.[96] Rule 34(b) states that a party wishing to inspect documents or things or to enter upon the property of another need only serve on the other party a request for permission to do so.[97] The other party then serves a response either granting the request or stating reasons for objection.[98]

Paragraph (c) of Rule 34 has been amended. The former Rule 34(c) stated that Rule 34 did not preclude an independent action against nonparties for the production of documents and things and for permission to enter upon land.[99] The rule now provides:

> (c) **Persons not Parties.** A person not a party to the action may be compelled to produce documents and things or to submit to an inspection as provided in Rule 45.[100]

The language of Rule 34(c) was changed to permit courts to compel nonparties to produce documents or submit to an inspection relating to an action.[101] The change in Rule 34(c) reflects changes made in Rule 45.[102] Rule 45 now provides for subpoenas to compel nonparties

---

91. *Id.*

92. *See* 8 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2202 (1970).

93. *See* Hickman v. Taylor, 329 U.S. 495, 507 (1947) (Mutual knowledge of the relevant facts gathered by both parties is essential to proper litigation and reduces the possibility of surprise.).

94. *See* WRIGHT & MILLER, *supra* note 92, § 2206.

95. *See id.*

96. FED. R. CIV. P. 34(a).

97. *See* FED. R. CIV. P. 34(b); *see also* WRIGHT & MILLER, *supra* note 92, § 2207.

98. FED. R. CIV. P. 34(b).

99. FED. R. CIV. P. 34(c), 28 U.S.C. app. 593 (1982).

100. FED. R. CIV. P. 45(c).

101. *See Amendments to Federal Rules Are Approved by Supreme Court*, 59 U.S.L.W. 2695 (1991).

102. *Id. See infra* notes 148-93 and accompanying text.

to produce documents and things and to submit to inspections of premises.[103]

The amendment to Rule 34(c) alters the outcome in cases such as *Fleming v. Gardner*.[104] In *Fleming* the defendants moved the court to force a nonparty hospital to produce documents and to allow the defendants to inspect and copy or photograph records of the treatment and hospitalization of the plaintiff.[105] The court refused to order the hospital to produce the documents because Rule 34 "creates a discovery device that may be used only against parties to a pending action."[106] Because the hospital was not a party to the action, the court could not issue the order under Rule 34.[107] Under the present Rule 34(c), the court clearly could have ordered the hospital to produce the records.[108]

Rule 35. Physical and Mental Examinations of Persons

Under Rule 35, whenever the physical or mental condition of a party is in controversy, the court in which the action is pending may require the party to submit to a physical or mental examination.[109] The court may order the examination upon a showing that the mental or physical condition of a party is "in controversy" and that there is "good cause" to order the examination.[110] To establish good cause for an examination, a party must make an affirmative showing that the condition as to which the examination is sought is, in fact, important to the controversy and that there is good reason for ordering the examination.[111] An order for the physical or mental examination of a party is not granted as of right but is a matter of discretion with the trial judge, who must determine whether the party seeking the order has met the "in controversy" and "good cause" requirements.[112]

---

103. *See* FED. R. CIV. P. 34 advisory committee's note.

104. 84 F.R.D. 217 (E.D. Tenn. 1978).

105. *Id.*

106. *Id.* (citing Hickman v. Taylor, 329 U.S. 495 (1945), and WRIGHT & MILLER, *supra* note 92, § 2208). *See also* Hilgenberg v. Neth, 93 F.R.D. 325, 326 (E.D. Tenn. 1981) (denying defendant's motion seeking orders directing certain nonparty corporations to produce desired documents).

107. *Fleming*, 84 F.R.D. at 217. The court did, however, order the plaintiff to acquire the documents from the hospital and produce them for inspection and copying by the defendants. *Id.* at 218.

108. *See* FED. R. CIV. P. 34(c).

109. *See* FED. R. CIV. P. 35; WRIGHT & MILLER, *supra* note 92, § 2231; *see also* Sibbach v. Wilson & Co., 312 U.S. 1, 14 (1941) (Rule 35 does not invade the substantive rights of a party and is consistent with the Rules Enabling Act.).

110. *See* FED. R. CIV. P. 35.

111. *See* Schlagenhauf v. Holder, 379 U.S. 104, 118 (1964).

112. *See* WRIGHT & MILLER, *supra* note 92, § 2234; *see also* Sanden v. Mayo Clinic, 495 F.2d 221, 225 (8th Cir. 1974) (Rule 35(a) motion rests in the sound discretion of the court); Bucher v. Krause, 200 F.2d 576, 584 (7th Cir. 1952) (same result).

Paragraph (a) of Rule 35 has been amended substantively while paragraph (b) has been amended only to reflect the language change in paragraph (a). The rule now reads:

> (a) **Order for Examination.** When the mental or physical condition (including the blood group) of a party or of a person in the custody or under the legal control of a party, is in controversy, the court in which the action is pending may order the party to submit to a physical or mental examination by a suitably licensed or certified examiner or to produce for examination the person in the party's custody or legal control. The order may be made only on motion for good cause shown and upon notice to the person to be examined and to all parties and shall specify the time, place, manner, conditions, and scope of the examination and the person or persons by whom it is to be made.
>
> (b) **Report of Examiner.**
>
> (1) If requested by the party against whom an order is made under Rule 35(a) or the person examined, the party causing the examination to be made shall deliver to the requesting party a copy of the detailed written report of the examiner setting out the examiner's findings, including results of all tests made, diagnoses and conclusions, together with like reports of all earlier examinations of the same condition. After delivery the party causing the examination shall be entitled upon request to receive from the party against whom the order is made a like report of any examination, previously or thereafter made, of the same condition, unless, in the case of a report of examination of a person not a party, the party shows that the party is unable to obtain it. The court on motion may make an order against a party requiring delivery of a report on such terms as are just, and if an examiner fails or refuses to make a report the court may exclude the examiner's testimony if offered at trial.
>
> * * * * *
>
> (3) This subdivision applies to examinations made by agreement of the parties, unless the agreement expressly provides otherwise. This subdivision does not preclude discovery of a report of an examiner or the taking of a deposition of the examiner in accordance with the provisions of any other rule.[113]

The amendments to Rule 35(a) widen the range of physical and mental examinations a court may order. Previously, the rule only allowed examinations by physicians.[114] In 1988, the rule was revised to allow mental examinations by licensed clinical psychologists.[115] The 1991 amendment authorizes the court to require physical or mental examinations conducted by any person "suitably licensed or certi-

---

113. FED. R. CIV. P. 35.

114. *See* WRIGHT & MILLER, *supra* note 92, § 2235 (former rule only refers to examinations by a physician).

115. *See* FED. R. CIV. P. 35 advisory committee's note.

fied."[116] This means a party may be compelled, with proper notice,[117] to submit to an examination by a dentist, an occupational therapist, or any other professional required to be licensed.[118] The examining person need only be qualified to give valuable testimony about the physical or mental condition in dispute.[119]

The requirement that an examiner should be licensed or certified to conduct the type of examination needed is designed to encourage courts to use discretion before ordering an examination.[120] The court should determine whether the proposed examiner is qualified to examine a party before any examination occurs.[121] This should assure that no person is subjected to a court-ordered examination by an examiner whose testimony would be of such limited value as to make it unjust to order the person to undergo the examination.[122] The court's discretion to order the examination should extend to physicians as well.[123] If the proposed examination calls for an expertise the physician does not have, the court should not order the examination.[124]

Rule 41. Dismissal of Actions

Rule 41 establishes methods of dismissing an action.[125] Dismissal may be voluntary under Rule 41(a) or involuntary under Rule 41(b).[126] The general purpose of Rule 41(a) is to preserve the plaintiff's right to take a voluntary nonsuit and start over so long as the defendant is not prejudiced.[127] The rule allows voluntary dismissals only before service of an answer or a motion for summary judgment in order to limit the right to dismissal to an early stage of

---

116. FED. R. CIV. P. 35(a).

117. Rule 35(a) retains the requirement that notice of the order for the examination must be given to the person to be examined and to all parties. *See* FED. R. CIV. P. 35(a); *see also* Liechty v. Terrill Trucking Co., 53 F.R.D. 590, 591 (E.D. Tenn. 1971) (notice must be sent to all parties, not just adverse counsel).

118. *See* FED. R. CIV. P. 35 advisory committee's note.

119. *Id.*

120. *Id. See supra* note 112 and accompanying text.

121. *See* FED. R. CIV. P. 35 advisory committee's note.

122. *See* Schlagenhauf v. Holder, 379 U.S. 104, 120-21 (1964) (An allegation that the driver of a bus involved in an accident had impaired vision does not justify an order that he submit to examinations in the areas of internal medicine, ophthalmology, neurology, and psychiatry.); *see also* WRIGHT & MILLER, *supra* note 92, at § 2234 (court may refuse to order examination or may restrict examination to the particular aspect of the party's health in question).

123. *See* FED. R. CIV. P. 35 advisory committee's note.

124. *Id.*

125. *See generally* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2363 (1971).

126. FED. R. CIV. P. 41.

127. *See* McCall-Bey v. Franzen, 777 F.2d 1178, 1184 (7th Cir. 1985).

the proceedings.[128] If an answer or a motion for summary judgment has been filed and all of the parties to the action did not stipulate to voluntary dismissal, a plaintiff who wishes to dismiss must obtain an order of the court,[129] which may be granted at the court's discretion.[130]

Rule 41(b) primarily provides defendants with the ability to ask a court to dismiss a plaintiff's case.[131] Paragraph (b) has been amended and now states:

> (b) **Involuntary Dismissal: Effect Thereof.** For failure of the plaintiff to prosecute or to comply with these rules or any order of court, a defendant may move for dismissal of an action or of any claim against the defendant. Unless the court in its order for dismissal otherwise specifies, a dismissal under this subdivision and any dismissal not provided for in this rule, other than a dismissal for lack of jurisdiction, for improper venue, or for failure to join a party under Rule 19, operates as an adjudication upon the merits.[132]

The amendment deleted the second and third sentences of the former Rule 41(b). These sentences had authorized the use of Rule 41(b) as a method of dismissing a nonjury action on the merits if the plaintiff failed to carry the burden of proof in presenting his or her case.[133] The old motion to dismiss under Rule 41(b) on grounds that the plaintiff's evidence is legally insufficient now should be pursued as a motion for judgment on partial findings authorized by the amended Rule 52(c).[134] The amended Rule 41(b) retains the directive that an involuntary dismissal pursuant to the rule is with

---

128. *See* D.C. Elecs. v. Nartron Corp., 511 F.2d 294, 296-97 (6th Cir. 1975). The dismissal is typically without prejudice unless a plaintiff's action based on the same claim has been dismissed from any court of the United States or any state. *See* FED. R. CIV. P. 41(a)(1).

129. FED. R. CIV. P. 41(a)(2); WRIGHT & MILLER, *supra* note 125, § 2364. Dismissal under Rule 41(a)(2) is also typically without prejudice. *See* FED. R. CIV. P. 41(a)(2).

130. *See* Garner v. Missouri-Pacific Lines, 409 F.2d 6, 7 (6th Cir. 1969); Stevenson v. United States, 197 F. Supp. 355, 357 (M.D. Tenn. 1961). *See generally* Lawrence Mentz, *Voluntary Dismissal by Order of Court-Federal Rules of Civil Procedure Rule 41(a)(2) and Judicial Discretion*, 48 NOTRE DAME L. REV. 446 (1972).

131. *See* WRIGHT AND MILLER, *supra* note 125, § 2369; *see also* Societe Internationale Pour Participations Industrielles et Commerciales S.A. v. Rogers, 357 U.S. 197, 207 (1958) (Rule 41(b), on its face, is appropriate only as a defendant's remedy.).

132. FED. R. CIV. P. 41(b).

133. *See* FED. R. CIV. P. 41 advisory committee's note.

134. *Id. See infra* notes 225-32 and accompanying text. Rule 52(c) now authorizes the entry of a judgment against the defendant as well as the plaintiff and allows it at an earlier time than the close of the case of the party against whom judgment has been rendered. *See infra* note 250-51 and accompanying text.

prejudice and operates as an adjudication on the merits.[135]

Rule 44. Proof of Official Record

Rule 44 provides a method of authenticating official records for use in an action. Generally, Rule 44 allows an official, or her agent, to certify the authenticity of documents within her custody.[136] The purpose of the rule is to avoid the burden, expense, and inconvenience of calling a public official to court for the purpose of testifying to the authenticity of a document.[137] While Rule 44(a)(1) provides the method of authenticating domestic records, Rule 44(a)(2) provides a method of authenticating foreign official records.[138]

Rule 44(a) has been amended to read:

> (a) **Authentication.**
>
> (1) Domestic. An official record kept within the United States, or any state, district, or commonwealth, or within a territory subject to the administrative or judicial jurisdiction of the United States, or an entry therein, when admissible for any purpose, may be evidenced by an official publication thereof or by a copy attested by the officer having the legal custody of the record, or by the officer's deputy, and accompanied by a certificate that such officer has the custody. The certificate may be made by a judge of a court of record of the district or political subdivision in which the record is kept, authenticated by the seal of the court, or may be made by any public officer having a seal of office and having official duties in the district or political subdivision in which the record is kept, authenticated by the seal of the officer's office.
>
> (2) Foreign. A foreign official record, or an entry therein, when admissible for any purpose, may be evidenced by an official publication thereof; or a copy thereof, attested by a person authorized to make the attestation, and accompanied by a final certification as to the genuineness of the signature and official position (i) of the attesting person, or (ii) of any foreign official whose certificate of genuineness of signature and official position relates to the attestation or is in a chain of certificates of genuineness of signature and official position relating to the attestation. A final certification may be made by a secretary of embassy or legation, consul general, vice consul, or consular agent of the United States, or a diplomatic or consular official of the foreign country assigned or accredited to the United States. If reasonable opportunity has been given to all parties to investigate the authenticity and accuracy of the documents, the court may, for good cause shown, (i) admit an attested copy

---

135. *See* FED. R. CIV. P. 41(b).

136. *See* FED. R. CIV. P. 44(a).

137. *See generally* Lester B. Orfield, *Proof of Official Records in Federal Cases*, 22 MONT. L. REV. 137 (1961).

138. *See* WRIGHT & MILLER, *supra* note 125, §§ 2434-35.

> without final certification or (ii) permit the foreign official record to be evidenced by an attested summary with or without a final certification. The final certification is unnecessary if the record and the attestation are certified as provided in a treaty or convention to which the United States and the foreign country in which the official record is located are parties.[139]

Rule 44(a)(1) has been modified only slightly. The new rule does not refer to specific territories, and the new rule uses only generic terms to describe those governments having a relationship with the United States such that their documents should be treated as domestic records.[140] The methods of authenticating a domestic record are the same as under the former rule. The record may be authenticated by official publication or by an attested copy accompanied by a certificate proving the attesting official had actual possession of the document.[141]

Rule 44(a)(2) as amended concerns the authentication of foreign official records and retains nearly all the language of the former rule. As under the former rule, the new rule allows foreign documents to be authenticated by official publication, certified attested copy, or by "chain certification."[142] The new rule also retains the former rule's provision that a court has the discretion to admit an attested copy of a document without certification or an attested summary of a record with or without a final certification.[143]

The drafters of the amendments added a new sentence at the end of Rule 44(a)(2). This sentence eliminates the need for final certification by United States diplomatic officials if the United States and the foreign country where the document is located have entered a treaty or agreement abolishing the requirement.[144] If this is the situation, the directives of the treaty or agreement should be fol-

---

139. FED. R. CIV. P. 44(a).

140. *See* FED. R. CIV. P. 44 advisory committee's note on 1991 amendment rules. The generic terms used are "a territory subject to the administrative or judicial jurisdiction of the United States." FED. R. CIV. P. 44(a)(1).

141. *Id.*

142. *See* FED. R. CIV. P. 44(a)(2). Chain certification involves a situation where the original attestation purports to have been made by an authorized person and is accompanied by a certificate of another foreign official whose certificate may, in turn, be followed by that of a foreign official of higher rank, and so on. *See* WRIGHT & MILLER, *supra* note 125, § 2435.

143. *See* FED. R. CIV. P. 44(a)(2). In order for the court to exercise its discretion, there must be "good cause shown," and all parties must have had a reasonable time to investigate the authenticity and accuracy of the documents. *Id.*

144. *Id.* The new sentence read, "The final certification is unnecessary if the record and attestation are certified as provided in a treaty or convention to which the United States and the foreign country in which the official record is located are parties. *Id.* This does not change the practice of attesting the records under the old rule; it merely changes the certification requirements for that attestation. *See* FED. R. CIV. P. 44 advisory committee's note on 1991 amendment rules.

lowed.[145] This amendment to Rule 44(a)(2) only applies to those countries that are parties to the Hague Convention Abolishing the Requirement of Legalisation for Public Documents.[146] For other nations the authentication procedure remains the same as under the former rule.[147]

Rule 45. Subpoena

Rule 45, which provides for the issuance of subpoenas,[148] has been rewritten completely. The Advisory Committee's goals in rewriting Rule 45 were (1) to give greater protection to those persons who are required to give information or evidence to a court; (2) to provide easier access to documents and other information in the possession of nonparties to an action; (3) to facilitate the service of subpoenas in districts other than the district where the action is proceeding; (4) to enable a court to compel a witness found within the state where court sits to attend trial; and (5) to better organize the text of the rule.[149]

The former rule provided, in part, that to issue a subpoena, one had to apply to the clerk of court for the subpoena.[150] The clerk would issue a subpoena captioned only out of the clerk's own court, leaving most of the required information "in blank."[151] The attorney would then fill in the information and arrange for service.[152] The rule required service by the marshal or by a nonparty at least eighteen years of age.[153] A subpoena for documents from nonparties had to

---

145. *Id.*

146. *See* FED. R. CIV. P. 44 advisory committee's note on 1991 amendment rules; *see also* Convention Abolishing the Requirement of Legalisation for Foreign Public Documents, October 5, 1961, 33 U.S.T. 883, 527 U.N.T.S. 189 (adopted by the United States on October 15, 1981). Currently, the following nations are parties to the Hague Public Documents Convention: Antigua and Barbuda, Argentina, Austria, Bahamas, Belgium, Botswana, Brunei, Cyprus, Fiji, Finland, France, Germany, Greece, Hungary, Israel, Italy, Japan, Lesotho, Liechtenstein, Luxembourg, Malawi, Malta, Mauritius, Netherlands, Norway, Portugal, Seychelles, Spain, Suriname, Swaziland, Switzerland, Tonga, Turkey, United Kingdom and Northern Ireland, United States, and Yugoslavia. FED. R. CIV. P. 44 tbl. (West Supp. 1991). *See generally* William C. Harvey, *The United States and the Hague Convention Abolishing the Requirement of Legalisation for Foreign Documents*, 11 HARV. INT'L L.J. 476 (1970).

147. *See* FED. R. CIV. P. 44(a)(2).

148. *See generally* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2451 (1971) (purpose and history of Rule 45).

149. FED. R. CIV. P. 45 advisory committee's note on 1991 amendment rules.

150. *See* Fed. R. Civ. P. 45(a), 28 U.S.C. app. 611 (1982).

151. *Id.*

152. *Id.*

153. *See* Fed. R. Civ. P. 45(c), 28 U.S.C. app. 611 (1982).

be served through a subpoena forcing the nonparty to attend a deposition.[154]

Rule 45 now reads:

(a) **Form; Issuance.**

(1) Every subpoena shall

(A) state the name of the court from which it is issued; and

(B) state the title of the action, the name of the court in which it is pending, and its civil action number; and

(C) command each person to whom it is directed to attend and give testimony or to produce and permit inspection and copying of designated books, documents or tangible things in the possession, custody or control of that person, or to permit inspection of premises, at a time and place therein specified; and

(D) set forth the text of subdivisions (c) and (d) of this rule.

A command to produce evidence or to permit inspection may be joined with a command to appear at trial or hearing or at deposition, or may be issued separately.

(2) A subpoena commanding attendance at a trial or hearing shall issue from the court for the district in which the hearing or trial is to be held. A subpoena for attendance at a deposition shall issue from the court for the district designated by the notice of deposition as the district in which the deposition is to be taken. If separate from a subpoena commanding the attendance of a person, a subpoena for production or inspection shall issue from the court for the district in which the production or inspection is to be made.

(3) The clerk shall issue a subpoena, signed but otherwise blank, to a party requesting it, who shall complete it before service. An attorney as an officer of the court may also issue and sign a subpoena on behalf of

(A) a court in which the attorney is authorized to practice; or

(B) a court for a district in which a deposition or production is compelled by the subpoena, if the deposition or production pertains to an action pending in a court in which the attorney is authorized to practice.

(b) **Service.**

(1) A subpoena may be served by any person who is not a party and is not less than 18 years of age. Service of a subpoena upon a person named therein shall be made by delivering a copy thereof to such person and, if the person's attendance is commanded, by tendering to that person the fees for one day's attendance and the mileage allowed by law. When the subpoena is issued on behalf of the United States or an officer or agency thereof, fees and mileage need not be tendered. Prior notice of any commanded production of documents and things or inspection of premises before trial shall be served on each party in the manner prescribed by Rule 5(b).

---

154. *See* Fed. R. Civ. P. 45(d), 28 U.S.C. app. 611-12 (1982).

(2) Subject to the provisions of clause (ii) of subparagraph (c)(3)(A) of this rule, a subpoena may be served at any place within the district of the court by which it is issued, or at any place without the district that is within 100 miles of the place of the deposition, hearing, trial, production, or inspection specified in the subpoena or at any place within the state where a state statute or rule of court permits service of a subpoena issued by a state court of general jurisdiction setting in the place of the deposition, hearing, trial, production, or inspection specified in the subpoena. When a statute of the United States provides therefor, the court upon proper application and cause shown may authorize the service of a subpoena at any other place. A subpoena directed to a witness in a foreign country who is a national or resident of the United States shall issue under the circumstances and in the manner and be served as provided in Title 28, U.S.C. § 1783.

(3) Proof of service when necessary shall be made by filing with the clerk of the court by which the subpoena is issued a statement of the date and manner of service and of the names of the persons served, certified by the person who made the service.

(c) **Protection of Persons Subject to Subpoenas.**

(1) A party or an attorney responsible for the issuance and service of a subpoena shall take reasonable steps to avoid imposing undue burden or expense on a person subject to that subpoena. The court on behalf of which the subpoena was issued shall enforce this duty and impose upon the party or attorney in breach of this duty an appropriate sanction, which may include, but is not limited to, lost earnings and a reasonable attorney's fee.

(2)(A) A person commanded to produce and permit inspection and copying of designated books, papers, documents or tangible things, or inspection of premises need not appear in person at the place of production or inspection unless commanded to appear for deposition, hearing or trial.

(B) Subject to paragraph (d)(2) of this rule, a person commanded to produce and permit inspection and copying may, within 14 days after service of the subpoena or before the time specified for compliance if such time is less than 14 days after service, serve upon the party or attorney designated in the subpoena written objection to inspection or copying of any or all of the designated materials or of the premises. If objection is made, the party serving the subpoena shall not be entitled to inspect and copy the materials or inspect the premises except pursuant to an order of the court by which the subpoena was issued. If objection has been made, the party serving the subpoena may, upon notice to the person commanded to produce, move at any time for an order to compel the production. Such an order to compel production shall protect any person who is not a party or an officer of a party from significant expense resulting from the inspection and copying commanded.

(3)(A) On timely motion, the court by which a subpoena was issued shall quash or modify the subpoena if it

(i) fails to allow reasonable time for compliance;

(ii) requires a person who is not a party or an officer of a

party to travel to a place more than 100 miles from the place where that person resides, is employed or regularly transacts business in person, except that, subject to the provisions of clause (c)(3)(B)(iii) of this rule, such a person may in order to attend trial be commanded to travel from any such place within the state in which the trial is held, or

(iii) requires disclosure of privileged or other protected matter and no exception or waiver applies, or

(iv) subjects a person to undue burden.

(B) If a subpoena

(i) requires disclosure of a trade secret or other confidential research, development, or commercial information, or

(ii) requires disclosure of an unretained expert's opinion or information not describing specific events or occurrences in dispute and resulting from the expert's study made not at the request of any party, or

(iii) requires a person who is not a party or an officer of a party to incur substantial expense to travel more than 100 miles to attend trial, the court may, to protect a person subject to or affected by the subpoena, quash or modify the subpoena or, if the party in whose behalf the subpoena is issued shows a substantial need for the testimony or material that cannot otherwise be met without undue hardship and assures that the person to whom the subpoena is addressed will be reasonably compensated, the court may order appearance or production only upon specified conditions.

(d) **Duties in Responding to Subpoena.**

(1) A person responding to a subpoena to produce documents shall produce them as they are kept in the usual course of business or shall organize and label them to correspond with the categories in the demand.

(2) When information subject to a subpoena is withheld on a claim that is privileged or subject to protection as trial preparation materials, the claim shall be made expressly and shall be supported by a description of the nature of the documents, communications, or things not produced that is sufficient to enable the demanding party to contest the claim.

(e) **Contempt.** Failure by any person without adequate excuse to obey a subpoena served upon that person may be deemed a contempt of the court from which the subpoena issued. An adequate cause for failure to obey exists when a subpoena purports to require a non-party to attend or produce at a place within the limits provided by clause (ii) of subparagraph (c)(3)(A).[155]

Subdivision (a) contains amendments to the former rule. First, paragraph (a)(1) sets forth the form every subpoena must take, *i.e.*, whether the subpoena is issued to command a witness to appear, to

155. FED. R. CIV. P. 45. *See* David D. Siegel, *Federal Subpoena Practice Under the New Rule 45 of the Federal Rules of Civil Procedure*, 139 F.R.D. 197, 199-202 (1992).

produce documents, or to permit inspections of a premises.[156] Second, paragraph (a)(1) authorizes the issuance of a subpoena to compel a nonparty to produce evidence without having to attend a deposition.[157] This amendment reflects the change to Rule 34 now permitting courts to force nonparties to produce evidence or allow inspections.[158] Third, paragraph (a)(2) allows an attorney to require a person subject to a subpoena to produce materials in that person's control whether or not the materials are located in the district where the subpoena can be served.[159] Fourth, and perhaps most importantly, paragraph (a)(3) authorizes attorneys to issue a subpoena on behalf of a court in which they are authorized to practice.[160] Rule 45 also authorizes attorneys to issue a subpoena on behalf of a court in a distant district to compel a deposition or the production of evidence needed for an action in a court in which the attorney is authorized to practice.[161] This amendment allows attorneys to act as public officers entitled to use the court's contempt power[162] to investigate facts in dispute, and it effectively authorizes service of a subpoena anywhere in the United States by an attorney representing any party.[163]

Subdivision (b), which provides for service of subpoenas, is a compilation of various sections from the former Rule 45 with a few minor changes. Paragraph (b)(1) retains the basic text of former subdivision (c), but the Advisory Committee deleted any reference to service by marshal or deputy marshal[164] simply because of the infrequent use of these officers.[165] Paragraph (b)(1) also requires that when a person receives prior notice of a command to produce documents or to allow inspection of property, the prior notice must be served on the parties to an action in the manner prescribed by Rule 5(b).[166] This notice allows all the parties to an action the opportunity to monitor discovery, object to discovery, or demand further discovery.[167] Paragraph (b)(2), which provides for where

---

156. FED. R. CIV. P. 45.

157. *See* FED. R. CIV. P. 45 advisory committee's notes.

158. *See supra* notes 92-108 and accompanying text.

159. *See* FED. R. CIV. P. 45 advisory committee's note.

160. *See* FED. R. CIV. P. 45 (a)(3)(A).

161. *See* FED. R. CIV. P. 45(a)(3)(B).

162. *See* FED. R. CIV. P. 45(e); *see also* Young v. United States *ex rel.* Vuitton et Fils S.A., 481 U.S. 787, 821 (1987) (Scalia, J., concurring in judgment) (Courts are empowered to prosecute for contempt "those who . . . disobey orders necessary for the conduct of [their] business (such as subpoenas).").

163. *See* FED. R. CIV. P. 45 advisory committee's note. The purpose of this amendment is to alleviate the administrative expense and delay of interdistrict practice. *Id.*

164. *See* FED. R. CIV. P. 45(b)(1).

165. *See* FED. R. CIV. P. 45 advisory committee's note.

166. *See* FED. R. CIV. P. 45 (b)(1); *see also* FED. R. CIV. P. 5(b).

167. *See* FED. R. CIV. P. 45 advisory committee's note.

subpoenas may be served, retains the language of former Rule 45's subdivision (e) and now applies not only to subpoenas for hearing and trials but also to depositions, productions, and inspections.[168] Paragraph (b)(3) retains some of the language of former Rule 45's subdivision (d)(1). Attorneys, however, must now file, when necessary, proof of service with the clerk of court for all subpoenas, not just those for depositions as provided by the former Rule 45.[169]

Subdivision (c) of rule 45 is almost completely new. The new rule sets the rights of persons subject to subpoenas and also outlines the liabilities of parties who misuse the subpoena power.[170] The Advisory Committee determined that this section was necessary because the greatly increased power of attorneys to issue and serve subpoenas could be subject to abuse.[171]

Paragraph (c)(1) begins by imposing an affirmative duty on parties or attorneys to avoid subjecting a person against whom a subpoena is issued to undue burden or expense.[172] This requirement merely reflects the duty imposed by Rule 26(g) not to pursue unreasonable or unduly burdensome discovery in light of the needs of a case.[173] Rule 45(c)(1) now gives a court the right to impose sanctions including, but not limited to, ordering the party or attorney to compensate the person against whom the subpoena is issued for lost earnings and reasonable attorney's fees resulting from abuse of the subpoena power.[174]

Paragraph (c)(2) extends the period available to object to the demands of a subpoena from ten days to fourteen days.[175] This amendment is designed to make calculations of time easier for Rule 6 purposes[176] and to allow more time for persons subject to a subpoena to object.[177] The rule also provides that if a court order is necessary to compel production the court order should protect any person who is not a party to the action from excessive expense involved with involuntary assistance to the court.[178]

---

168. *See* FED. R. CIV. P. 45(b)(2).

169. *See* FED. R. CIV. P. 45 advisory committee's note.

170. FED. R. CIV. P. 45(c).

171. FED. R. CIV. P. 45(c) advisory committee's note ("The liability of the attorney is correlative to the power of the attorney to issue subpoenas.").

172. FED. R. CIV. P. 45(c)(1); *Amendments to Federal Rules are Approved By Supreme Court*, 59 U.S.L.W. 2695 (1991).

173. *See* FED. R. CIV. P. 26(g)(13); *see also* MODEL RULES OF PROFESSIONAL CONDUCT RULE 4.4 (1990) ("A lawyer shall not use means that have no substantial purpose other than to embarrass, delay, or burden a third person.").

174. FED. R. CIV. P. 45(c)(1).

175. FED. R. CIV. P. 45(c)(2)(B).

176. FED. R. CIV. P. 45(c) advisory committee's note; *see also* FED. R. CIV. P. 6(a) ("When period of time prescribed is less than 11 days, intermediate Saturdays, Sundays, and legal holidays are excluded in the computation.").

177. FED. R. CIV. P. 45(c) advisory committee's note.

178. FED. R. CIV. P. 45(c)(2)(B).

Paragraph (c)(3) authorizes courts to quash subpoenas to protect persons from any misuse of the subpoena power.[179] The rule sets forth in detail the grounds upon which a subpoena can be quashed or modified.[180] One important change has been made. Under the new rule, a federal court can compel a witness to come from any place in the state to attend trial whether or not state law so provides.[181] This authority is subject to Rule 45(c)(3)(B)(iii), which allows a court to protect or assure compensation to a nonparty witness where there could be undue hardship.[182]

Subparagraph (c)(3)(B) provides that, in certain circumstances, a subpoena should be quashed unless the serving party shows substantial need and the court can protect the interests of the subpoenaed person. The rule allows a court to quash or modify a subpoena to protect persons subject to or affected by a subpoena from disclosing trade secrets, confidential research, or other intellectual property.[183] The intellectual property of both nonparty witnesses and experts is protected.[184] The clause is not intended to protect information or opinions of retained experts[185] but is meant to protect unretained experts who may nevertheless still be subject to subpoena.[186] The Advisory Committee describes the practice of compelling unretained experts to testify as a "growing problem" that threatens the intellectual property of these experts by denying them the opportunity to bargain for the value of their services.[187] As noted earlier the rule also protects nonparty witnesses who may have to travel substantial distances by allowing a court to condition a subpoena requiring travel more than 100 miles on the provision of reasonable compensation.[188]

Subdivision (d) of Rule 45 is new. Paragraph (d)(1) requires nonparties to produce and label documents demanded by subpoena as they are kept in the usual course of business or in the categories

---

179. FED. R. CIV. P. 45(c)(3).

180. *Id.*

181. FED. R. CIV. P. 45 advisory committee's note.

182. FED. R. CIV. P. 45(c)(3)(B)(iii).

183. FED. R. CIV. P. 45(c)(3)(B)(i). *See also* FED. R. CIV. P. 26(c)(7) (protective order can be issued to protect intellectual property).

184. FED. R. CIV. P. 45(c)(3)(B)(ii).

185. FED. R. CIV. P. 26(b)(4).

186. FED. R. CIV. P. 45(c)(3)(B)(ii) advisory committee's note.

187. *Id. See* Buchanan v. American Motors Corp., 697 F.2d 151, 152 (6th Cir. 1987) (no abuse of discretion for district court to quash subpoena requiring expert who was a stranger to the litigation to spend a large amount of time itemizing and explaining data to develop a research opinion). *See generally* Virginia G. Maurer, *Compelling the Expert Witness: Fairness and Utility Under the Federal Rules of Civil Procedure*, 19 GA. L. REV. 71 (1984); Mark Labaton, Note, *Discovery and Testimony of Unretained Experts: Creating a Clear and Equitable Standard to Govern Compliance with Subpoenas*, 1987 DUKE L.J. 140 (1987).

188. *See* FED. R. CIV. P. 45(c)(3)(B)(iii).

defined in the subpoena.[189] Paragraph (d)(2) requires that when a person refuses to produce information based on claims of privilege or work product, the refusing person must provide adequate information about the reasons for refusal so the discovering party can evaluate the refusal to determine whether it is justified.[190] A refusing person who fails to provide information about the privilege asserted is subject to an order to show cause why the person should not be held in contempt.[191]

Finally, subdivision (e) of Rule 45 states that persons who refuse to obey subpoenas may be held in contempt of court. Only persons who have an "adequate excuse" will avoid being held in contempt for refusal to obey a subpoena.[192] What exactly constitutes "adequate excuse" is undefined, but the rule does state that a person may not be held in contempt if the person is a nonparty and is compelled to travel greater distances than can be compelled pursuant to Rule 45.[193]

Rule 47. Selection of Jurors

Rule 47, providing for the selection of juries for federal actions, has undergone significant change under the 1991 amendments. Rule 47(a), which gives the court broad discretion in the conduct of the voir dire examination of prospective jurors,[194] remains unchanged. The former Rule 47(b), however, which allowed a court to impanel alternate jurors,[195] has been eliminated.

The use of alternate jurors is no longer needed in federal actions.[196] The former Rule 47(b) was based on the assumption that a jury should consist of exactly twelve members plus additional, or alternate, jurors in case original jurors were excused or disqualified.[197] Because twelve-member juries are no longer considered a constitu-

---

189. *See* FED. R. CIV. P. 45(d)(1); *see also* FED. R. CIV. P. 34(b) (Rule 45(d)(1) extends to nonparties the duty already imposed on parties to produce documents as kept in the usual course of business or as demanded).

190. *See* FED. R. CIV. P. 45(d)(2).

191. *See* FED. R. CIV. P. 45 advisory committee's note.

192. *See* FED. R. CIV. P. 45(e).

193. *Id.*

194. *See* Eisenhauer v. Burger, 431 F.2d 833, 836 (6th Cir. 1970) (questioning permitted during voir dire is in sound discretion of the court). *See generally* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2482 (1971).

195. *Id.* at § 2484. A federal court had authority to impanel alternate jurors to prevent mistrials in cases of long duration. If a regular juror died, became so ill that they could not continue their duties, or were disqualified, the alternate juror could take their place. *Id.*

196. *See* FED. R. CIV. P. 47 advisory committee's note ("[T]he institution of the alternate juror [is] abolished.").

197. *Id.*

tional necessity,[198] the use of alternate jurors is no longer required.

The remainder of Rule 47 now reads:

> (b) **Peremptory Challenges.** The court shall allow the number of peremptory challenges provided by 28 U.S.C. § 1870.
> (c) **Excuse.** The court may for good cause excuse a juror from service during trial or deliberation.[199]

Paragraphs (b) and (c) are new. Paragraph (b) directs the district court to only allow the number of peremptory challenges[200] for jurors as provided in 28 U.S.C. section 1870.[201] Title 28 U.S.C. section 1870[202] directs that in federal civil actions, each side must be permitted three peremptory challenges.[203] In single-party litigation, a court may not authorize more than three peremptory challenges per side,[204] while in multiple-party litigation, the court is authorized to use broad discretion in awarding peremptory challenges.[205]

Paragraph (c) is self-explanatory. The rule simply allows a court to dismiss a juror for good cause during a trial or even during the

---

198. Colgrove v. Battin, 413 U.S. 149, 157 (1973) ("It cannot be said that 12 members is a substantive aspect of the right of trial by jury."). Local rules of the United States District Courts for the Eastern and Middle Districts of Tennessee state that, except where required by law, the civil jury shall consist of six members and an appropriate number of alternates. *See* E.D. TENN. R. 48.1; M.D. TENN. R. 12(j). In light of the elimination of alternate jurors in federal court, it would be prudent to amend such local rules. *See* FED. R. CIV. P. 48 advisory committee's note (While juries may consist of six members, with the elimination of the alternate juror, it would be prudent and necessary, in order to provide for sickness or excused absence, to seat more than six jurors.). Local Rule 18 for the United States District Court for the Western District of Tennessee requires that civil juries consist of seven persons, but trials may continue with as few as six jurors. *See* W.D. TENN. R. 18.

199. FED. R. CIV. P. 47.

200. A peremptory challenge is "the right to challenge a juror without assigning, or being required to assign, a reason for the challenge." BLACK'S LAW DICTIONARY 1136 (6th ed. 1990).

201. 28 U.S.C. § 1870 (1988) provides in relevant part:

> In civil cases, each party shall be entitled to three peremptory challenges. Several defendants or several plaintiffs may be considered as a single party for the purposes of making challenges, or the court may allow additional peremptory challenges and permit them to be exercised separately or jointly.

202. *Id.*

203. *Id. E.g.*, Kotler v. American Tobacco Co., 926 F.2d 1217, 1226 (1st Cir. 1990).

204. Blount v. Plovidba, 567 F.2d 583, 585 (3d Cir. 1977).

205. 28 U.S.C. § 1870 (1988) (second sentence). *E.g.*, Fedorchick v. Massey-Ferguson, Inc., 577 F.2d 856, 858 (3d Cir. 1978) (statute permits trial judge to allocate more than three peremptory challenges when more than one plaintiff or defendant is present in a case). If a single party faces several parties with adverse interests, the trial judge may expand the number of peremptory challenges available to both sides. *See* Cary v. Lykes Bros. Steamship Co., 455 F.2d 1192, 1194 (5th Cir. 1972) (28 U.S.C. § 1870 permits trial judge to allocate ten peremptory challenges for plaintiff and five peremptory challenges apiece for two defendants).

jury deliberation stage of the trial without causing a mistrial.[206] Sickness, family emergency, and juror misconduct may constitute "good cause" to excuse a juror.[207]

Rule 48. Number of Jurors—Participation in Verdict

Rule 48 has been rewritten completely. The Advisory Committee determined that the former Rule 48, which allowed parties to stipulate that a jury could be composed of less than twelve persons, had become obsolete because of the adoption in many districts of local rules setting the standard size of a civil jury as six persons.[208] Rule 48 now states:

> The court shall seat a jury of not fewer than six and not more than twelve members and all jurors shall participate in the verdict unless excused from service by the court pursuant to Rule 47(c). Unless the parties otherwise stipulate, (1) the verdict shall be unanimous and (2) no verdict shall be taken from a jury reduced in size to fewer than six members.[209]

The rule expressly allows trial courts to seat juries with membership between six and twelve members.[210] The Advisory Committee, on the other hand, recommends that a trial court impanel more jurors than the minimum number of six in order to provide for the occurrence of sickness or disability among jurors.[211] This is especially important now because the use of alternate jurors has been abolished. If a court impanels more jurors than the minimum required, the possibility of mistrial is reduced because all seated jurors will have participated in the verdict; and if any jurors are excused through Rule 47,[212] the remaining jurors will be able to render a unanimous verdict of six or more.[213]

The new Rule 48 allows parties to stipulate to nonunanimous verdicts and verdicts from juries with less than six members,[214] but

---

206. FED. R. CIV. P. 47(c).

207. FED. R. CIV. P. 47 advisory committee's note.

208. FED. R. CIV. P. 48 advisory committee's note. *See, e.g.*, E.D. TENN. R. 48.1; M.D. TENN. R. 12(j).

209. FED. R. CIV. P. 48.

210. *Id.* The rule follows the opinion of the United States Supreme Court that a jury with fewer that six jurors may be inconsistent with the Seventh Amendment. *See, e.g.*, Ballew v. Georgia, 435 U.S. 223, 239 (1978); *cf.* Hanson v. Parkside Surgery Ctr., 872 F.2d 745 (6th Cir. 1989) (submitting case to eight member jury was not constitutional error).

211. *See* FED. R. CIV. P. 48 advisory committee's note.

212. Rule 47 allows a court to excuse jurors for "good cause." FED. R. CIV. P. 47. *See supra* notes 178-93 and accompanying text.

213. *See* FED. R. CIV. P. 48 advisory committee's note.

214. *See* FED. R. CIV. P. 48.

the Advisory Committee does not recommend making such stipulations.[215] Parties should only agree to a smaller jury in "exceptional circumstances" because of the constitutional importance of the right to a jury trial and because "smaller juries are more erratic and less effective in serving to distribute responsibility for the exercise of judicial power."[216] Furthermore, parties should only agree upon nonunanimous verdicts in exceptional circumstances where the jury unexpectedly drops in number.[217]

Rule 50. Judgment as a Matter of Law in Actions Tried by Jury; Alternative Motion for New Trial; Conditional Rulings

While Rule 50 has undergone significant change, the rule still allows a court to take away from the consideration of the jury cases in which the facts are sufficiently clear that the law requires a particular result.[218] The former Rule 50 provided for a motion for "directed verdict" at the close of the evidence before a case was submitted to a jury.[219] The motion allowed a court to determine whether there were any questions of fact and whether a jury possibly could reach a verdict different from the one directed.[220] If a jury reached a verdict the court felt was erroneous as a matter of law or unsupported by evidence, the former Rule 50 allowed a court to set aside the verdict of the jury and enter a "judgment notwithstanding the verdict."[221]

The new Rule 50 has eliminated the terms "directed verdict" and "judgment notwithstanding the verdict."[222] Rule 50 now states:

> (a) **Judgment as a Matter of Law.**
>
> (1) If during a trial by jury a party has been fully heard with respect to an issue and there is no legally sufficient evidentiary basis for a reasonable jury to have found for that party with respect to that issue, the court may grant a motion for judgment as a matter of law against that party on any claim, counterclaim, cross-claim, or third party claim that cannot under the controlling law be maintained without a favorable finding on that issue.

---

215. *See* FED. R. CIV. P. 48 advisory committee's note.

216. *Id.*

217. *Id.* The Advisory Committee warns courts that they should not rely on agreements to accept nonunanimous verdicts because of the constitutional difficulties with juries of less than six members. *Id.*

218. *See generally* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2521 (1971) (history and purpose of former Rule 50).

219. *Id.*

220. *Id.* The rule was designed to save the time and trouble involved in lengthy jury determinations. *Id.* *See* Martin v. Erie-Lackawanna R.R., 388 F.2d 802, 804-05 (6th Cir. 1968).

221. *See* WRIGHT AND MILLER, *supra* note 218.

222. *See* FED. R. CIV. P. 50.

(2) Motions for judgment as a matter of law may be made at any time before submission of the case to the jury. Such a motion shall specify the judgment sought and the law and the facts on which the moving party is entitled to the judgment.

(b) **Renewal of Motion for Judgment after Trial; Alternative Motion for New Trial.** Whenever a motion for a judgment as a matter of law made at the close of all the evidence is denied or for any reason is not granted, the court is deemed to have submitted the action to the jury subject to a later determination of the legal questions raised by the motion. Such a motion may be renewed by service and filing not later than 10 days after entry of the judgment. A motion for a new trial under Rule 59 may be joined with a renewal of the motion for judgment as a matter of law, or a new trial may be requested in the alternative. If a verdict was returned, the court may, in disposing of the renewed motion, allow the judgment to stand or may reopen the judgment and either order a new trial or direct the entry of judgment as a matter of law. If no verdict was returned, the court may, in disposing of the renewed motion, direct the entry of judgment as a matter of law or may order a new trial.

(c) **Same: Conditional Rulings on Grant of Motion for Judgment as a Matter of Law.**

(1) If the renewed motion for judgment as a matter of law is granted, the court shall also rule on the motion for a new trial, if any, by determining whether it should be granted if the judgment is thereafter vacated or reversed, and shall specify the grounds for granting or denying the motion for the new trial. If the motion for a new trial is thus conditionally granted, the order thereon does not affect the finality of the judgment. In case the motion for a new trial has been conditionally granted and the judgment is reversed on appeal, the new trial shall proceed unless the appellate court has otherwise ordered. In case the motion for a new trial has been conditionally denied, the appellee on appeal may assert error in that denial; and if the judgment is reversed on appeal, subsequent proceedings shall be in accordance with the order of the appellate court.

(2) The party against whom judgment as a matter of law has been rendered may serve a motion for a new trial pursuant to Rule 59 not later than 10 days after entry of the judgment.

(d) **Same: Denial of Motion for Judgment as a Matter of Law.** If the motion for judgment as a matter of law is denied, the party who prevailed on that motion may, as appellee, assert grounds entitling the party to a new trial in the event the appellate court concludes that the trial court erred in denying the motion for judgment. If the appellate court reverses the judgment, nothing in this rule precludes it from determining that the appellee is entitled to a new trial, or from directing the trial court to determine whether a new trial shall be granted.[223]

---

223. FED. R. CIV. P. 50.

Rule 50 now uses the term "judgment as a matter of law" to describe what used to be both the directed verdict and the judgment notwithstanding the verdict.[224] The motion for a directed verdict is now known as the motion for judgment as a matter of law while the motion for judgment notwithstanding the verdict is simply a "renewed" motion for judgment as a matter of law.[225] The Advisory Committee notes that, in essence, the former motions for directed verdict and judgment notwithstanding the verdict were the same in character but were made at different times in the proceeding.[226] The new terminology eliminates a meaningless distinction.

Rule 50 has been amended in other respects. Paragraph (a) now articulates the standard for when judgment as a matter of law should be granted.[227] The motion should be granted when there is no evidence sufficient for a reasonable jury to find for a party on an issue.[228] This standard is not new; however, it has been found previously only in case law.[229] The rule also allows a court to enter judgment as a matter of law at any time during the trial before the case has been submitted to the jury.[230] The court, however, must be certain that a party has been "fully heard with respect to an issue" before judgment as a matter of law may be granted.[231]

Rule 50(a)(2) has retained the requirement that in order for a party to reserve its right to move for a renewed motion for judgment as a matter of law after a jury has returned a verdict, a motion for judgment as a matter of law must be made at the close of all the evidence.[232] This requirement serves the important purpose of allowing the nonmoving party to cure any deficiencies in proof that may have gone unnoticed before a jury begins deliberations.[233] If a party is

---

224. *Id.*

225. *See* FED. R. CIV. P. 50(b).

226. *See* FED. R. CIV. P. 50 advisory committee's note (directed verdict and judgment notwithstanding the verdict had "common identity"); *see also* O'Neill v. Kiledjian, 511 F.2d 511, 513 (6th Cir. 1975) (standard for measuring the legal sufficiency of directed verdict and judgment notwithstanding the verdict is the same).

227. *See* FED. R. CIV. P. 50(a).

228. *Id.*

229. *E.g.*, Hale v. Holy Cross Hosp. Inc., 513 F.2d 315, 318 n.2 (5th Cir. 1975); *see* Edward H. Cooper, *Directions for Directed Verdicts: A Compass for Federal Courts*, 55 MINN. L. REV. 903 (1971).

230. *See* FED. R. CIV. P. 50(a).

231. *Id.* The Advisory Committee states that a court should not enter judgment against a party when the party has not been warned of the importance of a dispositive fact at issue and has not been allowed to present all available evidence bearing on that fact. *See* FED. R. CIV. P. 50 advisory committee's note.

232. *See* FED. R. CIV. P. 50(a)(2).

233. *See* FED. R. CIV. P. 50 advisory committee's note; *see also* Benson v. Allphin, 786 F.2d 268, 273 (7th Cir. 1986) ("[A] motion for directed verdict at the close of all the evidence provides the nonmovant with an opportunity to do what he can to remedy the deficiencies in his case" before the jury retires to deliberate.).

allowed to move for a post-verdict judgment as a matter of law after not moving for judgment as a matter of law at the close of evidence, the opposing party may be prejudiced by having lost the opportunity to present additional evidence regarding an issue before the case is submitted to a jury.[234]

Paragraph (a)(2) now also requires that the motion for judgment as a matter of law clearly state the judgment sought and the grounds on which judgment should be granted.[235] This requirement serves the purpose of giving notice to the responding party that there may be deficiencies in proof and giving the party the opportunity to correct those deficiencies. The requirement is also designed to end the practice in some courts of allowing post-verdict motions for judgment as a matter of law even when the moving attorney made no formal motion for judgment at the close of evidence.[236] By requiring specificity in the judgment sought and the facts and law in support of that judgment, this practice is now prohibited.[237]

Paragraph (b) of Rule 50 has retained most of the substance of the old rule. A post-verdict motion for judgment as a matter of law is simply a renewal of the pre-verdict motion.[238] As such, a post-trial motion for judgment can only be granted on the grounds set forth in the pre-verdict motion.[239] In addition, the new rule retains the requirement that a post-trial motion for judgment as a matter of law must be served and filed not later than ten days after entry of judgment.[240] Paragraphs (c) and (d) of Rule 50 have been amended to reflect the change in terminology in Paragraphs (a) and (b).

Rule 52. Findings by the Court; Judgment on Partial Findings

Rule 52 directs a court to make findings of fact and conclusions of law in all actions tried upon the facts without a jury or advisory jury and in granting or denying interlocutory injunctions.[241] The

---

234. *See* Farley Transp. Co. v. Santa Fe Trail Transp. Co., 786 F.2d 1342, 1346 (9th Cir. 1985) ("[F]ailure to renew an earlier motion for a directed verdict may lull the opposing party into believing that the moving party has abandoned any challenge to the sufficiency of the evidence.").

235. *See* FED. R. CIV. P. 50(a)(2).

236. *E.g.*, Villanueva v. McInnis, 723 F.2d 414 (5th Cir. 1984); *see also* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2537 (1971) (courts take liberal view on what constitutes a motion for directed verdict in deciding whether there was a sufficient prerequisite for the motion for judgment notwithstanding the verdict).

237. *See Amendments to Federal Rules are Approved by Supreme Court*, 59 U.S.L.W. 2695, 2696 (1991).

238. *See* WRIGHT & MILLER, *supra* note 236.

239. *Id. E.g.*, Smith v. Lightning Bolt Prods. Inc., 861 F.2d 363, 367 (2d Cir. 1988).

240. FED. R. CIV. P. 50(b).

241. *See* WRIGHT & MILLER, *supra* note 236, § 2571.

language of the former Rule 52(a) has been altered to eliminate any reference to the former Rule 41(b) in the last sentence of the rule.[242] The former Rule 52(a) required courts to make findings of fact and conclusions of law when dismissal was sought under the former Rule 41(b), which allowed dismissal when the plaintiff had failed to carry an essential burden of proof.[243] The remainder of the amended rule retains the language of the former Rule 52(a), but a new paragraph (c) has been added to the rule.

The rule now provides:

> (a) **Effect.** In all actions tried upon the facts without a jury or with an advisory jury, the court shall find the facts specially and state separately its conclusions of law thereon, and judgment shall be entered pursuant to Rule 58; and in granting or refusing interlocutory injunctions the court shall similarly set forth the findings of fact and conclusions of law which constitute the grounds of its action. Requests for findings are not necessary for purposes of review. Findings of fact, whether based on oral or documentary evidence, shall not be set aside unless clearly erroneous, and due regard shall be given to the opportunity of the trial court to judge of the credibility of the witnesses. The findings of a master, to the extent that the court adopts them, shall be considered as the findings of the court. It will be sufficient if the findings of fact and conclusions of law are stated orally and recorded in open court following the close of the evidence or appear in an opinion or memorandum of decision filed by the court. Findings of fact and conclusions of law are unnecessary on decisions of motions under Rule 12 or 56 or any other motion except as provided in subdivision (c) of this rule.
>
> * * * * *
>
> (c) **Judgment on Partial Findings.** If during a trial without a jury a party has been fully heard with respect to an issue and the court finds against the party on that issue, the court may enter judgment as a matter of law against that party on any claim, counterclaim, cross-claim or third-party claim that cannot under the controlling law be maintained or defeated without a favorable finding on that issue, or the court may decline to render any judgment until the close of all the evidence. Such a judgment shall be supported by findings of fact and conclusions of law as required by subdivision (a) of this rule.[244]

Paragraph (c) allows a court to render "judgment on partial findings" in bench trials.[245] "At any time during trial after a party has been fully heard on an issue, the court . . . may enter judgment as a matter of law against the party on any claim that cannot be

---

242. FED. R. CIV. P. 52(a), 28 U.S.C. app. 617 (1982).
243. *Id.*
244. FED. R. CIV. P. 52.
245. *Id.*

maintained under the controlling law absent a favorable finding on that issue."[246] The new paragraph (c) replaces a part of Rule 41(b) deleted by the 1991 amendments,[247] which allowed dismissal at the close of the plaintiff's case if the plaintiff failed to carry an essential burden of proof.[248] This explains the deletion in the final sentence of paragraph (a) of Rule 52 of any mention of Rule 41(b).[249] Under the new Rule 52(c) any party, plaintiff or defendant, may move for judgment on partial findings.[250] Entry of judgment on the motion is in the sound discretion of the court and may be entered during trial if a party has been fully heard on the issue or after the close of evidence.[251]

Rule 53. Masters

Rule 53 allows federal district courts to appoint masters.[252] Masters are appointed to assist the court in obtaining facts and to help the court reach a correct result.[253] The use of masters is reserved primarily for cases involving complex issues of fact and law "so technical or esoteric as to be outside ordinary judicial competence."[254] Rule 53 has been amended only slightly. Paragraph (e) now provides:

> (e) **Report.**
>
> (1) Contents and Filing. The master shall prepare a report upon the matters submitted to the master by order of reference and, if required to make findings of fact and conclusions of law, the master shall set them forth in the report. The master shall file the report with the clerk of court and serve on all parties notice of the filing. In an action to be tried without a jury, unless otherwise directed by the order of reference, the master shall file with the report a transcript of the proceedings and of the evidence and the original exhibits. Unless otherwise directed by the order of reference, the master shall serve a copy of the report on each party.[255]

Paragraph (e) retains the requirement of the former Rule 53 that masters prepare a report on the issues presented to them and that

---

246. *See Amendments to Federal Rules are Approved by Supreme Court*, 59 U.S.L.W. 2695, 2696 (1991).

247. *See supra* notes 125-35 and accompanying text.

248. *See* FED. R. CIV. P. 52 advisory committee's note.

249. *Id.*

250. *See* FED. R. CIV. P. 52 advisory committee's note.

251. FED. R. CIV. P. 52(c).

252. *See* 9 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2601 (1971).

253. *Id.*

254. *See* Note, *Masters and Magistrates in the Federal Courts*, 88 HARV. L. REV. 779, 795 (1975).

255. FED. R. CIV. P. 53(e).

they file the report with the clerk of court.[256] The rule has also been amended to require the master to serve on all parties notice that the report has been filed plus a copy of the report itself.[257] The former rule only required a filing of the report, with the clerk notifying the parties of the filing.[258] A party then had to obtain a copy from the clerk.[259] The amendment is designed to ameliorate proceedings requiring a master and to save counsel and clerks of court time and effort.[260]

Rule 63. Inability of Judge to Proceed

Rule 63 states the procedure if a judge, for whatever purpose, is unable to proceed with a hearing or trial that is in progress.[261] The rule has been rewritten and now states:

> If a trial or hearing has been commenced and the judge is unable to proceed, any other judge may proceed with it upon certifying familiarity with the record and determining that the proceedings in the case may be completed without prejudice to the parties. In a hearing or trial without a jury, the successor judge shall at the request of a party recall any witness whose testimony is material and disputed and who is available to testify again without undue burden. The successor judge may also recall any other witness.[262]

The new Rule 63 is significantly different from the old. The former rule made no provision for disqualification or other withdrawal of a judge and was limited to the death, sickness, or disability of the presiding judge.[263] The amended rule allows for substitution of a different judge regardless of the reason for the original judge's withdrawal.[264] The former rule also provided only for withdrawal of the judge after the trial and made no provision for substitution during trial.[265] Rule 63 now allows cases to proceed before a successor

---

256. *Id.*
257. *Id.*
258. *See* FED. R. CIV. P. 53 advisory committee's note.
259. *Id.*
260. *Id.*
261. *See generally* 11 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2921 (1973).
262. FED. R. CIV. P. 63.
263. *See* FED. R. CIV. P. 63 advisory committee's note.
264. *See* FED. R. CIV. P. 63. The Advisory Committee states that under the amended rule judges may discontinue only for compelling reasons, which should be stated on the record. FED. R. CIV. P. 63 advisory committee's note.
265. *See* Morton v. Ortho Pharmaceutical Corp., 550 F. Supp. 416, 417 (E.D. Tenn. 1982) (judge died suddenly after trial had ended but before jury was charged); Comment, *The Case of the Dead Judge: Fed. R. Civ. P. 63:* Whalen v. Ford Motor Credit Co., 67 MINN. L. REV. 827 (1983).

judge so long as the new judge can certify familiarity with the record and determine that the case can be completed before that judge without prejudice to the parties.[266] Needless to say, the requirement that the judge be familiar with the record is designed to avoid any unfairness that may result from a judge's unfamiliarity with an action.[267] The availability of a transcript or a videotape of the trial proceedings prior to substitution is necessary for the judge's certification.[268]

The amended rule allows a successor judge to recall witnesses where needed so long as the recall is not overly burdensome.[269] For nonjury trials Rule 63 also allows a successor judge to make findings of fact based upon the evidence heard by the original judge.[270] The successor judge may make findings of fact based on testimony that the judge deems undisputed;[271] and, if a witness is unavailable, the successor judge may make a finding of fact based on testimony recorded at trial.[272] The Advisory Committee warns, however, that the court risks error by determining the credibility of a witness who has not been seen or heard but who is available for recall.[273]

Rule 72. Magistrates, Pretrial Matters

Rule 72 is designed to implement the legislative mandate of 28 U.S.C. section 636(b)(1)(A), which allows a federal district judge to designate a magistrate to hear and determine nearly all pretrial matters pending before the court.[274] Rule 72 divides the civil pretrial

---

266. *See* FED. R. CIV. P. 63 advisory committee's note; *cf.* FED. R. CRIM. P. 25(a) (in criminal proceeding successor judge must certify familiarity with the record to proceed with and finish trial).

267. *See* FED. R. CIV. P. 63 advisory committee's note.

268. *Id.* The Advisory Committee states that the prompt availability of the transcript or videotape is extremely important for the successful use of Rule 63 because a jury cannot be held for long amounts of time while a transcript is prepared without prejudice resulting for one party or another. *Id.*

269. *See* FED. R. CIV. P. 63.

270. *See* FED. R. CIV. P. 63 advisory committee's note.

271. *Id.*

272. *Id.* The recorded testimony is the equivalent of a recorded deposition for Rule 32 purposes. *Id.*

273. *Id.*

274. 28 U.S.C. § 636(b)(1)(A) (1988) states:

(b)(1) Notwithstanding any provision of law to the contrary—

(A) a judge may designate a magistrate to hear and determine any pretrial matter pending before the court, except a motion for injunctive relief, for judgment on the pleadings, for summary judgment, to dismiss or quash an indictment or information made by the defendant, to suppress evidence in a criminal case, to dismiss or to permit maintenance of a class action, to dismiss for failure to state a claim upon which relief can be granted, and to involuntarily dismiss an action. A judge of the court may reconsider any pretrial matter under this subparagraph (A) where it has been shown that the magistrate's order is clearly erroneous or contrary to law.

authority of magistrates into two categories.[275] Rule 72(a) describes the responsibilities of the magistrate for pretrial matters "not dispositive of a claim or defense of a party" and directs how objections to a magistrate's report should be made.[276] Rule 72(b) serves the same purpose for pretrial matters "dispositive of a claim or defense of a party . . . ."[277]

Rule 72(a) has been amended and now reads:

> (a) **Nondispositive Matters.** A magistrate to whom a pretrial matter not dispositive of a claim or defense of a party is referred to hear and determine shall promptly conduct such proceedings as are required and when appropriate enter into the record a written order setting forth the disposition of the matter. Within 10 days after being served with a copy of the magistrate's order, a party may serve and file objections to the order; a party may not thereafter assign as error a defect in the magistrate's order to which objection was not timely made. The district judge to whom the case is assigned shall consider such objections and shall modify or set aside any portion of the magistrates order found to be clearly erroneous or contrary to law.[278]

The language of the former Rule 77(c) has been amended only slightly. Similar to the former rule, the amended rule allows a magistrate to hear and make determinations of nondispositive pretrial matters and to enter into the record a written order describing the disposition of the matter.[279] Also like the former rule, the amended rule allows the district judge to whom the case is assigned to modify or set aside any part of the magistrate's order he or she finds to be clearly erroneous or contrary to law.[280]

The former rule has been amended to eliminate a discrepancy between paragraphs (a) and (b) of Rule 72. Under the former Rule 72, paragraph (a) required objections to the magistrate's handling of nondispositive matters to be served and filed within ten days of entry of the magistrate's order.[281] Paragraph (b) of Rule 72, however, directs that objections to the magistrate's handling of dispositive matters were to be made within ten days of being served with a copy of the recommended disposition.[282] Rule 72(a) has been amended to remove this discrepancy. Now, in both nondispositive and dispositive pretrial matters handled by a magistrate, parties must object within

---

275. *See* FED. R. CIV. P. 72. *See generally* 12 CHARLES A. WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 3076.3 (Supp. 1991).

276. *Id. See* FED. R. CIV. P. 72(a).

277. *See* FED. R. CIV. P. 72(b).

278. FED. R. CIV. P. 72(a).

279. *See* FED. R. CIV. P. 77(d).

280. *Id.*

281. *See* FED. R. CIV. P. 72 advisory committee's note.

282. *Id.*

ten days of being served with a copy of the magistrate's recommended disposition.[283] The rule also has been amended to state clearly that untimely objections to a magistrate's order will not be considered.[284]

Rule 77. District Courts and Clerks

Rule 77(d) provides that the clerk of court shall serve a notice of entry of an order or judgment as provided in Rule 5 on each party who is not in default for failure to appear.[285] The clerk is to make a note in the docket of the mailing.[286] Notification by the clerk is simply for the convenience of the litigants.[287]

Rule 77(d) has been amended to provide:

> (d) **Notice of Orders or Judgments.** Immediately upon the entry of an order or judgment the clerk shall serve a notice of the entry by mail in the manner provided for in Rule 5 upon each party who is not in default for failure to appear, and shall make a note in the docket of the mailing. Any party may in addition serve a notice of such entry in the manner provided in Rule 5 for the service of papers. Lack of notice of the entry by the clerk does not affect the time to appeal or relieve or authorize the court to relieve a party for failure to appeal within the time allowed, except as permitted in Rule 4(a) of the Federal Rules of Appellate Procedure.[288]

The rule as amended is not materially different from the earlier version. As in the former rule, the amended rule provides that clerks shall serve orders or judgments on the parties to an action and shall note the mailing in the docket.[289] The amended rule also retains the former rule's direction that a clerk's failure to give notice of entry of an order or judgment does not increase the time allowed for appeal except as provided in Rule 4(a) of the Federal Rules of Appellate Procedure.[290]

Sentence two of the former rule began by stating that the mailing by the clerk of a notice of entry of judgment or order was sufficient notice for all purposes for which entry of an order is required.[291] This language has been deleted from the current rule.[292] The purpose of the amendment is to lessen the "draconian effect" of Rule 77(d)

---

283. FED. R. CIV. P. 72. *See also* M.D. TENN. R. 505.
284. *See* FED. R. CIV. P. 72(a).
285. *See* FED. R. CIV. P. 77(d).
286. *Id.*
287. *See* FED. R. CIV. P. 77 advisory committee's note to the 1946 amendment.
288. FED. R. CIV. P. 77(d).
289. *Id.*
290. *Id.*
291. *Id.*
292. FED. R. CIV. P. 77(d); *Amendments to Federal Rules are Approved by Supreme Court*, 59 U.S.L.W. 2695, 2696 (1991).

emerging in the federal courts.[293] In many circumstances parties against whom a judgment was entered failed to receive notice from the clerk of court of entry of judgment and lost the right to appeal the judgment because of a late filing of the notice of appeal.[294] Courts were imposing on counsel the duty to keep in almost constant contact with the clerk while a case was under consideration.[295] This "due diligence standard"[296] could be difficult to maintain for both counsel and court, especially if counsel was from outside the district.[297]

The amended rule relaxes the due diligence and harsh results imposed by the courts through Rule 77(d) in three ways. First, the amendment allows a district court, on motion, to extend the time to appeal in cases where a party's notice of appeal is filed late because of the party's failure to receive notice of the judgment against it.[298] The amended rule works in conjunction with Federal Rule of Appellate Procedure 4(a) to allow a district court to enlarge the time for appeal by fourteen days when the court finds that notice of the judgment was in fact not received by the losing party and that no prejudice would be caused by allowing the appeal.[299] Second, the somewhat relaxed language of the current rule is designed to encourage courts to be lenient on parties who have received no notice of entry of judgment.[300] Some courts have allowed parties who have

---

293. *See* FED. R. CIV. P. 77 advisory committee's note.

294. *See, e.g.*, Tucker v. Commonwealth Land Title Ins. Co., 800 F.2d 1054, 1055-56 (11th Cir. 1986) (A plaintiff who never received notice from clerk that adverse judgment had been entered could not appeal judgment four months after entry of judgment and three months after FED. R. APP. P. 4(a) thirty-day limit for filing notice of appeal had expired.).

295. *See* Wilson v. Atwood Group, 725 F.2d 255, 258 (5th Cir. 1984) (court may grant relief from thirty-day appeal limit when counsel who filed a belated appeal had been diligent in either attempting to delay the entry of judgment or in inquiring about the status of the case); *see also* FED. R. CIV. P. 77 advisory committee's note; Benjamin Calkins, *The Emerging Due Diligence Standard for Filing Delayed Notice of Appeal in Federal Courts*, 19 WILLAMETTE L. REV. 609 (1983).

296. *Id.*

297. *See* FED. R. CIV. P. 77 advisory committee's note.

298. *See Changes to Federal Practice Take Effect; Civil Procedure*, MASS. LAW. WKLY., Dec. 2, 1991.

299. FED. R. APP. P. 4(a)(6) provides:

> (6) The district court, if it finds (a) that a party entitled to notice of the entry of a judgment or order did not receive such notice from the clerk or any party within 21 days of its entry and (b) that no party would be prejudiced, may, upon motion filed within 180 days of entry of the judgment or order or within 7 days of receipt of such notice, whichever is earlier, reopen the time for appeal for a period of 14 days from the date of entry of the order reopening the time for appeal.

300. *See* FED. R. CIV. P. 77 advisory committee's note.

received no notice and who are outside the thirty-day appeals limit[301] to file a motion under Rule 60(b)(6)[302] requesting the court to vacate its final order and then reenter its judgment so that a timely notice of appeal can be filed.[303] Third, the amendment places the burden of notice on prevailing parties to assure that adversaries are aware of the entry of judgment.[304] If a prevailing party wants to be sure that the time for appeal is running, the prevailing party should be sure that opposing counsel has proper notice.[305]

CONCLUSION

Practitioners may expect more changes in the Federal Rules of Civil Procedure in 1992. The Supreme Court withheld from Congress proposed amendments to Rules 4, 4.1, 12, 26, 28, 30, and 71A for "further consideration" at the time the 1991 amendments were transmitted to Congress on April 30, 1991.[306] The proposed amendments to Rule 4 would substantially revamp the current methods of service of process by providing, among other things, that (1) the need for actual service could be eliminated with the assent of a defendant; (2) if a party fails to effect service on all officers required by law the court should allow a reasonable time to cure; and (3) a court can impose jurisdiction over the person of all defendants against whom federal law claims are made and who can be subject to the jurisdiction of the courts of the United States.[307] A number of new amendments to the Federal Rules of Civil Procedure have also been passed from the Advisory Committee to the Standing Committee for review and public comment.[308] Practitioners should be aware of the proposed amendments, comment on them if possible,[309] and may

---

301. *See* FED. R. APP. P. 4(a)(1).

302. Rule 60(b)(6) grants district courts the discretion to relieve parties from judgment or orders for "any . . . reason justifying relief from the operation of the judgment." FED. R. CIV. P. 60(b)(6).

303. *See* Expeditions Unlimited Aquatic Enters. v. Smithsonian Inst., 500 F.2d 808 (D.C. Cir. 1974).

304. *See* FED. R. CIV. P. 77 advisory committee's note.

305. *Id.*

306. *See* 134 F.R.D. 526 (April 30, 1991).

307. *See* FED. R. CIV. P. 4 advisory committee's note on proposed rule.

308. *See* 137 F.R.D. 63 (June 13, 1991). The amendments before the standing committee would affect Rules 1, 11, 16, 26, 29, 30, 31, 32, 33, 34, 36, 37, 43, 54, 56, 58, 83, and 84. *See id.* at 64-73.

309. Comments and suggestions on any of the federal rules can be sent to:
James E. Macklin, Jr.
Secretary
Committee on Rules of Practice and Procedure
Administrative Office of the United States Courts
Washington, D.C. 20544

expect at least some of these amendments to become effective December 1, 1992.

LANE MATTHEWS

RULES AND RIGHTS COLLIDING: SPEECH CODES AND THE FIRST AMENDMENT ON COLLEGE CAMPUSES

I. INTRODUCTION

Colleges and universities are being confronted increasingly with the use of "hate speech" on their campuses.[1] At the same time these institutions are under pressure to show sensitivity to the learning experiences of the divergent groups that make up the university community. In addressing these concerns, many colleges and universities have adopted speech codes. These codes are very controversial because they potentially sanction only "politically correct" language and prohibit legitimate, albeit controversial, speech.

The university's role in American society as a bastion of inquiry and thought, open to all ideas no matter how contentious or repugnant to accepted norms and established principles, is being questioned. The implementation of speech codes threatens academic freedom and, at public colleges and universities, gives rise to serious First Amendment issues.

II. BACKGROUND

In recent years, college and university campuses have been the scenes of increased use of hate speech by students.[2] At the University of Michigan, students at the campus radio station made racial jokes over the air.[3] At the University of Wisconsin, Zeta Beta Tau fraternity members dressed in blackface and participated in a slave auction.[4] At the University of Connecticut, Asian-American students were called "oriental faggots" by a group of football players.[5] At Arizona

---

1. Professor Nadine Strossen defines hate speech as "speech that expresses hatred or bias toward members of racial, religious, or other groups." Nadine Strossen, *Regulating Racist Speech on Campus: A Modest Proposal?*, 1990 DUKE L.J. 484, 487 [hereinafter *A Modest Proposal*].

2. *See* Charles R. Lawrence III, *If He Hollers Let Him Go: Regulating Racist Speech on Campus*, 1990 DUKE L.J. 431, 434; *A Modest Proposal*, *supra* note 1, at 487.

3. Jon Wiener, *Words That Wound: Free Speech for Campus Bigots*, 250 NATION, Feb. 26, 1990, at 272 [hereinafter *Words That Wound*].

4. Ken Emerson, *Only Correct—A Campus Report: Wisconsin; New Race Relation Rules on College Campuses,* 204 NEW REPUBLIC, Feb. 18, 1991, at 18 [hereinafter *Only Correct*].

5. *Words That Wound*, *see supra* note 3, at 272.

State University, students surrounded a black student and chanted, among other things, "Coon. Nigger. Porchmonkey."[6] These are only a few, illustrative examples of countless other incidents that are occurring.

The rise in these occurrences must be attributed in some respect to the increasingly diverse college community. At one time most college and university campuses contained fairly homogeneous student bodies. This is no longer true, though—universities are now multiracial, multicultural, and multiethnic institutions.[7] This change in community composition breeds an atmosphere in which the potential exists for ignorance and hatred to rear their ugly heads.

The rise in incidents of "hate speech" has coincided with a call for an educational experience that is sensitive to all groups within the student body. This effort is in response to incidents in which students and professors allegedly have offended particular groups on the college campus. To prevent this offensiveness many call for the avoidance of all references to "racist, sexist and homophobic ideas, attitudes and language."[8] For example, students at the District of Columbia School of Law demanded that a law professor stop representing a Georgetown student who was critical of Georgetown's affirmative action program.[9] Students at Harvard asked a law professor to refrain from discussing the admissibility of a woman's past sexual history at a rape trial because two students in his class were rape victims.[10] After a Scandinavian studies professor interpreted a novel in a manner that some found insensitive to women, students at the University of Minnesota accused the professor of sexual harassment.[11]

Under this framework, colleges and universities have adopted "speech codes" in an attempt to prevent the use of offensive speech. The University of Wisconsin's code provides that students may be sanctioned for "racist or discriminatory comments, epithets or other expressive behavior directed at an individual or on separate occasions at different individuals," and it then provides a laundry list of groups protected under the code.[12] Similar codes have been enacted at many

---

6. Jon Weiner, *Reagan's Children: Racial Hatred on Campus*, 248 NATION, Feb. 27, 1989, at 260 [hereinafter *Reagan's Children*].

7. The University of California at Berkeley is a good example of this development. Berkeley's incoming class in 1987 was 12% black, 17% Latino, 26% Asian, and 40% white. *Reagan's Children*, *supra* note 6, at 264.

8. Alan M. Dershowitz, *Politically Correct Is Intellectually Wrong*, NEWSDAY, Apr. 24, 1991, at 89.

9. *Warning This Editorial Is Not Politically Correct*, NEWSDAY, May 19, 1991, at 35.

10. *Id.*

11. *Id.*

12. *See Only Correct*, *supra* note 4, at 18.

universities both public and private.[13] Efforts by colleges to prevent all offensive speech, however, raise the questions of whether the codes stifle intellectual freedom and, at public universities, whether the codes trample upon First Amendment rights.

As a general proposition, the protections of the First Amendment extend to the public university setting as they do to any other setting in society.[14] The First Amendment, however, does not protect all speech and expression, although most speech and expression do fall within the scope of its protection.

III. THE SCOPE OF PROTECTED SPEECH

The First Amendment guarantees freedom of speech, but this freedom is not absolute. "[T]he lewd and obscene, the profane the libelous, and the insulting or 'fighting' words—those which by their very utterance inflict injury or tend to incite an immediate breach of the peace"[15] are not protected under the First Amendment. Society's interest in "order and morality" outweighs any interest that society has in protecting these types of "utterances."[16] As Justice Roberts stated in *Cantwell v. Connecticut*,[17] "Resort to epithets or personal abuse is not in any proper sense communication of information or opinion safeguarded by the Constitution . . . ."[18] This inexact rule is problematic because it is difficult to know with particularity what speech is protected since one person's "vulgarity is another's lyric."[19]

In one leading case, *Chaplinsky v. New Hampshire*,[20] the Court attempted to delineate that class of unprotected speech known as "fighting words."[21] Such words are defined as those "which by their very utterance . . . tend to incite an immediate breach of the peace."[22] Writing for the majority, Justice Murphy stated that "such utterances

---

13. "At the University of Pennsylvania students may be punished for using any language that 'stigmatizes or victimizes individuals' and 'creates an intimidating or offensive environment.'" *Words That Wound*, *supra* note 3, at 273. At the University of Connecticut, punishable offenses include the use of "derogatory names, inappropriately directed laughter, inconsiderate jokes, and conspicuous exclusion [of another student] from conversation." *Id.* According to a recent Boston Globe article, more than two hundred colleges and universities, both public and private, have enacted speech codes. *Speech Codes and Free Speech*, BOSTON GLOBE, Feb. 26, 1991, at 14.

14. *See infra* notes 57-68 and accompanying text.

15. Beauharnais v. Illinois, 343 U.S. 250, 256 (1952).

16. *Id.* at 257.

17. 310 U.S. 296 (1940).

18. *Id.* at 309-10.

19. Cohen v. California, 403 U.S. 15, 25 (1971).

20. 315 U.S. 568 (1942).

21. The words in question were "Your [sic] are a God damned racketeer" and "a damned Fascist." *Id.* at 569.

22. *Id.* at 572.

are no essential part of any exposition of ideas and [have] slight social value. . . .''[23] In determining whether words are fighting words, the majority set forth an objective standard to be used: Would a reasonable person be provoked to violence by the utterance of the offending words?[24]

In a later case, *Beauharnais v. Illinois*,[25] the Supreme Court appeared to expand the scope of unprotected speech when it upheld the conviction of a man for "unlawfully . . . exhibit[ing] in public places lithographs, which publications portray depravity, criminality, unchastity or lack of virtue of citizens of Negro race and color and which exposes [sic] citizens of Illinois of the Negro race and color to contempt, derision, or obloquy . . . .''[26] More specifically, the speech in question was a flyer in which Beauharnais blamed the community's problems on the "depravity" of its black members.[27] Beauharnais challenged his conviction on the grounds that it violated his First Amendment rights. The Court, however, upheld the conviction and labeled the speech libelous and unprotected under the Constitution.[28]

The Court defended the Illinois statute at issue by stating that the history of racial and religious prejudice in Illinois, and in the United States as a whole, provided the state a justifiable interest in seeking to prevent the intimidation and harassment of ethnic, religious, and racial groups.[29] The Court noted that the benefits from the

---

23. *Id.* (quoting *Cantwell*, 310 U.S. at 309-10).

24. *See Chaplinsky*, 315 U.S. at 573.

25. 343 U.S. 250 (1952).

26. *Beauharnais*, 343 U.S. at 252. The Court quoted the statute under which Beauharnais was convicted, which provided:

> It shall be unlawful for any person, firm or corporation to manufacture, sell, or offer for sale, advertise or publish, present or exhibit in any public place in this state any lithograph, moving picture, play, drama or sketch, which publication or exhibition portrays depravity, criminality, unchastity, or lack of virtue of a class of citizens, of any race, color, creed or religion which said publication or exhibition exposes the citizens of any race, color, creed, or religion to contempt, derision, or obloquy or which is productive of breach of the peace or riots.

Id. (quoting ILL. REV. STAT. ch. 38, § 471 (1949)).

27. Beauharnais posted lithographs that set forth a petition,

> [T]o halt the further encroachment, harassment and invasion of white people, their property, neighborhoods and persons, by the Negro. . . . If persuasion and the need to prevent the white race from becoming mongrelized by the negro will not unite us, then the aggressions . . . rapes, robberies, knives, guns and marijuana of the negro, surely will.

Beauharnais, 343 U.S. at 252.

28. *See id.* at 266.

29. *See id.* at 260-61. In justifying its decision the Court noted with approval the following commentary:

> [The words in question] are not the daily grist of vituperative political

speech in question were *de minimis*—adding little to public discourse—yet the harm to the objects of the speech was substantial.[30]

Beauharnais is especially noteworthy because the Court not only recognized the state's ability to prosecute a person for directing libelous words at an individual but also extended the state's power by allowing the prosecution of an individual for libeling a group. Writing for the majority, Justice Frankfurter stated:

> [G]roup-protection on behalf of the individual may . . . [be needed] in effectuating rights abstractly recognized as belonging to its members. . . . [W]e are precluded from saying that speech concededly punishable when immediately directed at individuals cannot be outlawed if directed at groups with whose position and esteem in society the affiliated individual may be inextricably involved.[31]

The only added requirement the Court seemed to place on "group libel" was that it be "calculated to have a powerful emotional impact on those to whom it was presented."[32] This group libel exception to the First Amendment has significant implications. If taken to the extreme, it would allow the prosecution of an individual for making any generalized statement about a group because generalizations by their nature are not categorically true.

Many have questioned this contraction of the First Amendment through the doctrine of "group libel." In fact, some members of the Supreme Court have questioned whether *Beauharnais* is still good law.[33] In a concurring opinion in *Garrison v. Louisiana*,[34] Justice Douglas questioned the continued validity of the case in light of Supreme Court cases that have greatly restricted individual libel claims.[35] In addition, several circuits have suggested what the Seventh Circuit has stated explicitly—Supreme Court libel cases "ha[ve] so washed away the foundations of *Beauharnais* that it could not be considered authoritative."[36]

> debate. Nor do they represent the frothy imaginings of lunatics, or the "idle" gossip of a country town. Rather, they indicate the systematic avalanche of falsehoods which are circulated concerning the various groups, classes and races which make up the countries of the western world.

Id. at 261 n.16 (quoting David Riesman, *Democracy and Defamation; Control of Group Libel*, 42 COLUM. L. REV. 727, 727 (1942)).

30. *See id.* at 256-57.

31. *Id.*, 343 U.S. at 262-63.

32. *Id.* at 261.

33. *See, e.g.*, Collin v. Smith, 578 F.2d 1197 (7th Cir. 1978).

34. 379 U.S. 64, 82 (1964) (Douglas, J., dissenting).

35. *See, e.g.*, New York Times v. Sullivan, 376 U.S. 254 (1964). This landmark case limited greatly a state's power to describe and prosecute libel. This limitation was based on concerns about chilling speech protected under the First Amendment.

36. American Booksellers Ass'n v. Hudnut, 771 F.2d 323, 331 n.3 (7th Cir.

In another important First Amendment case, *Brandenburg v. Ohio*,[37] the Supreme Court defined further the scope of speech that is not constitutionally protected.[38] Brandenburg was convicted under the Ohio Criminal Syndicalism Act for advocating violence against both African- and Jewish-Americans.[39] Brandenburg's advocacy was also peppered with racial epithets. The Court held the First Amendment protected Brandenburg's speech.[40] The Court did not view the words uttered by Brandenburg as "fighting words" but rather mere advocacy.[41] The Court held it is only when advocacy reaches a point at which there is a likelihood of persons being incited "to imminent lawless action" that speech loses its First Amendment protection.[42]

This holding refines the definition of unprotected speech. Apparently both "fighting words" and advocacy must incite imminent lawlessness in order to lose their First Amendment protection.[43] Imminent lawlessness has been interpreted by the Court to require a showing of "clear and present danger."[44] The implication of this test is that violence-producing speech is unprotected only when inciting "imminent . . . violence [that] cannot be satisfactorily prevented by means of crowd control techniques."[45]

In still another important First Amendment case, *Cohen v. California*,[46] the Court further refined the scope of the First Amendment protections.[47] In a seeming expansion of protected speech under the First Amendment, the Court held there is constitutional right to wear a jacket that has the message "Fuck the Draft" printed on it.[48] The majority rejected the notion that these words were obscene or that these were "fighting words," despite the fact that most in society would find the speech offensive.

In its opinion, the Court first considered the contention that the words in question were obscene. Justice Harlan, writing for the

---

1985), *aff'd*, 475 U.S. 1001 (1986); *see also* Dworkin v. Hustler Magazine Inc., 867 F.2d 1188, 1200 (9th Cir. 1989), *cert. denied*, 110 S. Ct. 59 (1989); *Collin*, 578 F.2d at 1205 (the appellate court doubting "that *Beauharnais* remains good law at all after the constitutional libel cases").

37. 395 U.S. 444 (1969).

38. *Id.*

39. More specifically, Brandenburg was convicted for "'advocat[ing] . . . the duty, necessity, or propriety of crime, sabotage, violence, or unlawful methods of terrorism as a means of accomplishing industrial or political reform' and for 'voluntarily asssembl[ing] with any society group, or assemblage of persons formed to teach or advocate the doctrines of criminal syndicalism.'" *Id.* at 445.

40. *Brandenburg*, 395 U.S. at 448-49.

41. *Id.*

42. *Id.*

43. *Id.* at 449.

44. *See id.* at 450-51 (Douglas, J., concurring).

45. LAWRENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW 855 (2d ed. 1988).

46. 403 U.S. 15 (1971).

47. *Id.*

48. *Id.* at 26.

majority, stated that in order for an expression to be considered obscene, the "expression must be, in some significant way erotic."[49] Because the words in question did not invoke any "psychic stimulation," the majority moved on to the question of whether these were "fighting words."[50]

Justice Harlan next rejected the idea that these were "fighting words." These three words, according to the majority, are not "inherently likely to provoke violent reaction" when "addressed to the ordinary citizen."[51] It is not enough that these words were "thrust" upon unwilling listeners, "the mere . . . presence of unwitting listeners . . . does not serve automatically to justify curtailing all speech capable of giving offense."[52] Only upon a showing that "substantial privacy interests are being invaded in an essentially intolerable manner" will speech lose its First Amendment protection.[53] The Court recognized this rule fosters a society in which "verbal tumult, discord, and even offensive utterance" are protected, but, according to the Court, this is preferable to having the government dictate what topics are open to public discourse.[54]

It is only through public discourse that "a more capable citizenry and more perfect polity" is created; furthermore, it is upon this "citizenry and polity" that the future of our democratic society rests.[55] "That the air may at times seem filled with verbal cacophony is, in this sense not a sign of weakness but of strength."[56]

Even though it is sometimes difficult to predict how the Court will apply the First Amendment to particular situations, the Court has been fairly consistent in holding that states may not regulate with impunity the content of speech. The Court in *Cohen* makes explicit what is implicit in many First Amendment decisions: The usual rule is that "governmental bodies may not prescribe the form or content of individual expression."[57]

IV. THE FIRST AMENDMENT ON COLLEGE AND UNIVERSITY CAMPUSES

Having established that the First Amendment does not protect all speech, the inquiry now focuses on the extent to which the First

---

49. *Id.* at 20; *see also* Roth v. United States, 354 U.S. 476 (1957).

50. *Cohen*, 403 U.S. at 20.

51. *Id.*

52. *Id.* at 21; *see, e.g.*, Organization for a Better Austin v. Keefe, 402 U.S. 415 (1971).

53. *Cohen*, 403 U.S. at 21.

54. *Id.* at 24-25.

55. *Id.* at 24; *see* Whitney v. California, 274 U.S. 357, 375-77 (1927) (Brandeis, J., concurring).

56. *Cohen*, 403 U.S. at 25.

57. *Id.* at 24. The Court also recognized there were established exceptions to this general principle. *See id.*

Amendment protects the speech of college and university students and teachers. This inquiry centers on public colleges and universities because private institutions are generally not subject to First Amendment constraints.[58]

In the preeminent case on the matter, *Tinker v. Des Moines Independent Community School District*,[59] the Court made clear that students and teachers do not "shed their constitutional rights to freedom of speech or expression at the schoolhouse gate."[60] Moreover, "[s]chool officials do not possess absolute authority over their students. Students in school as well as out of school are 'persons' under our constitution."[61] These First Amendment rights, however, must be tempered against the legitimate interest that schools have in maintaining order and discipline.[62]

Justice Fortas described the state's interest more explicitly, stating "conduct by the student, in class or out of it, which for any reason—whether it stems from time, place, or type of behavior—materially disrupts classwork or involves substantial disorder or invasion of the rights of others is . . . not immunized by the constitutional guarantee of freedom of speech."[63] This standard, however, varies somewhat based upon the type of school involved. The elementary school student's interest in free speech is different from that of the high school student, which is different from that of the college or university student.[64] Different maturity levels with the concomitant need for varying degrees of disciplinary control provide the justification for such distinctions.

The basic standard, however, applicable to all publicly funded schools, is that a school "must be able to show that its [prohibition on speech] was caused by something more than a mere desire to avoid the discomfort and unpleasantness that always accompany an unpopular viewpoint."[65] This standard is the result of the importance

---

58. Some in Congress would like to extend First Amendment protections to the private university setting. Representative Henry Hyde, a Republican from Illinois, has introduced the Collegiate Speech Protection Act of 1991. This bill would amend the U.S. Civil Rights Act of 1964 to extend First Amendment protections to private universities. *See* H.R. 1380, 102d Cong., 1st Sess. (1991); *see also* S. 1484, 102d Cong., 1st Sess. (1991) (Freedom of Speech on Campus Act of 1991).

59. 393 U.S. 503 (1969).

60. *Id.* at 506.

61. *Id.* at 511.

62. *See id.*

63. *Id.* at 513. This reflects generally accepted First Amendment jurisprudence, which rejects content-based restrictions on speech and only approves of those restrictions that affect the time, place, and manner of speech. *See supra* note 57 and accompanying text.

64. *Compare* Papish v. Board of Curators, 410 U.S. 667 (1973) (colleges) *with* Hazelwood Sch. Dist. v. Kuhlmeier, 484 U.S. 260 (1988) (high schools).

65. *Tinker*, 393 U.S at 509.

that the Supreme Court places on insuring that the First Amendment is part of the curriculum of public learning institutions. The Court views schools as vital to perpetuating this country's constitutional ideals and principles. Justice Brennan spoke eloquently on this subject in *Keyishian v. Board of Regents*.[66] He stated:

> The vigilant protection of constitutional freedoms is nowhere more vital than in the community of American schools. . . . The Nation's future depends upon leaders trained through wide exposure to that robust exchange of ideas which discovers truth "out of a multitude of tongues, [rather] than through any kind of authoritative selection."[67]

It is also worth noting that First Amendment protections not only extend to students but also to their teachers and professors as well. Chief Justice Warren, in *Sweezy v. New Hampshire*,[68] stated "Scholarship cannot flourish in an atmosphere of suspicion and distrust. Teachers and students must always remain free to inquire, to study and to evaluate, to gain new maturity and understanding; otherwise our civilization will stagnate and die."[69]

It would certainly be anomalous to deny students' First Amendment protections in an arena in which the exchange of ideas and quest for truth demands truly free speech. Furthermore, in preparing for the "real world" it is important for students to learn to cope with speech that they disagree with or find distasteful.

V. THE COURT'S COMMITMENT TO FREE SPEECH ON COLLEGE CAMPUSES

Having established that, in general, free speech is guaranteed on the college campus as well as having established the general limits of the First Amendment, the analysis now turns to specific cases in which public learning institutions have prohibited speech.

The Court has been faced with many cases concerning First Amendment rights on the college campus. In *Healy v. James*,[70] the Central Connecticut State College denied campus recognition to a group associated with militant activities directed at campus communities throughout the country.[71] The Court stated a "'heavy burden' rests on the college to demonstrate the appropriateness of [its] action."[72] Relying on *Brandenburg v. Ohio*,[73] the Court again distin-

66. 385 U.S. 589 (1967).
67. Keyishian v. Board of Regents, 385 U.S. 589, 603 (1967).
68. 354 U.S. 234 (1957).
69. Sweezy v. New Hampshire, 354 U.S. 234, 250 (1957).
70. 408 U.S. 169 (1972).
71. *Id.* at 174.
72. *Id.* at 184.
73. 395 U.S. 444 (1969); *see supra* notes 37-45 and accompanying text.

guished between action and mere advocacy. It was not enough that this group might advocate violence or even actually did advocate violence. Instead the group must actually "disrupt the work and discipline of the school" or interfere with the rights of others before the speech may be squelched.[74] The Court, however, did approve of "reasonable regulation with respect to the time, the place, and the manner in which student groups conduct their speech-related activities . . . ."[75] The Court has espoused this standard in other First Amendment cases outside of the academic setting.[76]

After *Healy v. James*, the Court was faced with the issue whether a university could expel a student for distributing a school newspaper that contained an article titled, "Mother Fucker Acquitted," which discussed an organization called "Up Against the Wall, Mother Fucker."[77] This edition of the newspaper also included a political cartoon "depicting policemen raping the Statue of Liberty and the Goddess of Justice."[78] The lower courts found the speech was unprotected under the First Amendment and upheld the student's expulsion.[79] The Supreme Court reversed.[80]

Noting that "state colleges and universities are not enclaves immune from the sweep of the First Amendment,"[81] the Court found that because the university had disapproved of the "*content* of the newspaper rather than the time, place, or manner of its distribution" there had been a violation of the student's First Amendment rights.[82] The Court viewed the speech as "the mere dissemination of ideas—no matter how offensive to good taste . . . ."[83] In a *per curiam* opinion the Court stated, "[T]he First Amendment leaves no room for the operation of a dual standard in the academic community with respect to the content of speech . . . ."[84] This is one of the strongest, if not *the* strongest, statements by the Court for protection of First Amendment rights on the college campus.

Chief Justice Burger's dissent, however, rejected this "extension" of the First Amendment, finding it "curious" and "bizarre" that the First Amendment prevented a state university from prohibiting the speech in question.[85] In fact, Chief Justice Burger thought that

---

74. *Healy*, 408 U.S. at 189.
75. *Id.* at 192-93.
76. *See, e.g., Cohen*, 403 U.S. at 15.
77. *See Papish*, 410 U.S. at 668.
78. *See id.* at 667.
79. *Id.* at 669.
80. *Id.* at 671.
81. *Id.* at 670 (quoting *Healy*, 408 U.S. at 180).
82. *Id.* (emphasis in original).
83. *Papish*, 410 U.S. at 670.
84. *Id.* at 671.
85. *Id.* at 672 (Burger, C.J., dissenting).

the majority's holding "demean[ed]" First Amendment values.[86] Furthermore, he stressed that part of the university's mission was to teach students to "express themselves in acceptable, civil terms."[87]

In *Tinker v. Des Moines Independent Community School District*,[88] the Court again took a strong stand for students' First Amendment rights.[89] Although this case involved a high school rather than a college, the language of the opinion applies with at least equal, if not greater, force to the college setting. In this case, students in Des Moines were disciplined for wearing black armbands in a protest of the Vietnam War.[90] The Court upheld the students' right to wear the armbands[91] and stated the school's "fear or apprehension of disturbance is not enough to overcome the right to freedom of expression. . . . Any word spoken . . . that deviates from the views of another person may start an argument or cause a disturbance. But our Constitution says we must take this risk."[92] This is true even though, as the Court recognized, the wearing of the armbands engendered some "hostile remarks" from other students.[93]

In the most recent case decided by the Court involving colleges and the First Amendment, *Widmar v. Vincent*,[94] the Court again upheld the sanctity of the amendment on the college campus.[95] The Court rejected as unconstitutional a public university's refusal to allow certain student groups to use campus facilities based on the views that the groups espoused.[96] The Court stated that in determining which groups could use the school for meetings, the school could not discriminate between groups based upon the content of a particular group's views.[97]

In decisions after *Widmar v. Vincent*, the Court has tended to be much less liberal in its application of First Amendment rights in the educational setting.[98] The Court is apparently headed in the direction that Chief Justice Burger, dissenting, urged in *Papish v. Board of Curators*.[99]

---

86. *Id.*
87. *Id.*
88. 393 U.S. 503 (1969).
89. *See id.* at 506.
90. *Id.* at 504.
91. *Id.*
92. *Tinker*, 393 U.S. at 508 (citing Terminiello v. Chicago, 337 U.S. 1 (1949)).
93. *Id.* at 508. The Court, however, was quick to note that no violence erupted. *Id.*
94. 454 U.S. 263 (1981).
95. *Id.* at 276.
96. *Id.* at 273.
97. *Id.* at 274-75.
98. *See, e.g.*, Papish v. Board of Curators, 410 U.S. 667 (1973).
99. *Id.*

Bethel School District No. 403 v. Fraser[100] illustrates this trend. Fraser, in a speech nominating a student at his high school for office, "referred to his candidate in terms of an elaborate, graphic, and explicit sexual metaphor."[101] The Court held that this language was unprotected by the First Amendment.[102] The Court distinguished this speech from the speech that occurred in *Tinker*[103] and found the language here to have no political value but instead to be "lewd and indecent."[104]

The Court reaffirmed its holdings recognizing that public sponsored education should instill the values of the First Amendment in the educational process.[105] Chief Justice Burger, however, added a gloss to the Court's previously strong protection of First Amendment rights at public learning institutions. The Chief Justice essentially incorporated his dissent in *Papish v. Board of Curators* into *Bethel*.[106] He stressed that not only did our "fundamental values" include those contained in the Constitution but also include "habits of civility."[107] He stated more specifically, "Even the most heated political discourse in a democratic society requires consideration for the personal sensibilities of the other participants and audiences."[108] The Court noted with approval the statement of the Second Circuit that "the First Amendment gives a high school student the classroom right to wear Tinker's armband, but not Cohen's jacket."[109] The importance of teaching civility seems now to rank very near the

---

100. 478 U.S. 675 (1986).

101. *Id.* at 677.

102. *See id.* at 682.

103. *Id.* at 680; *see also supra* notes 59-65 and accompanying text.

104. *Id.* at 685.

105. *Id.* at 681. The Court noted with approval the statement that public education's goal should be the "inculcat[ion of] fundamental values necessary to the maintenance of a democratic political system." Ambach v. Norwick, 441 U.S. 68, 76-77 (1979).

106. *See Bethel*, 478 U.S. at 681; *see also Papish*, 410 U.S. at 672.

107. *See Bethel*, 478 U.S. at 681.

108. *Id.* In a case decided after *Bethel*, the Court again called into question its previous strong position concerning the application of the First Amendment to public learning institutions. Though not characterizing it in terms of civility, the Court in *Hazelwood Sch. Dist. v. Kuhlmeier* limited the scope of the First Amendment in the public school setting. Students, who were members of the school newspaper staff, sued their principal and school district after the principal rejected as unsuitable an article the students had prepared for publication. *Hazelwood*, 484 U.S, at 264. One article dealt with teenage pregnancy, while the other dealt with the difficulties of dealing with the divorce of one's parents. *Id.* at 263. The Court held this speech was not protected by the First Amendment because the school had an interest in not embarrassing the students mentioned anonymously in the article. *Id.* at 274-76.

109. *Id.* at 682 (quoting Thomas v. Board of Educ., 607 F.2d 1043, 1057 (2d Cir. 1979), *cert. denied*, 444 U.S. 1081 (1980)).

importance that society has attached to instilling First Amendment values.

It is unclear whether this gloss applies with equal force to the university setting. The Court noted warily that very strong language was given protection in *Cohen v. California*[110] and distinguished that case from the instant case on the grounds that adults were involved.[111] Given this decision, the central issue facing colleges and universities attempting to regulate speech appears now to be: Are college students generally considered adults who enjoy the First Amendment protections of *Cohen v. California*,[112] or are college students first and foremost students who are controlled by *Bethel School District No. 403 v. Fraser* and therefore subject to learning the "habits and manners of civility?"[113] If the former, a dual standard of free speech emerges for high school students and college students.[114] If the latter, a dual standard of free speech rights emerges for college students and the rest of society. Whatever the case may be, it seems clear that dual standards now apply in some form to First Amendment rights in the academic community.[115]

Bethel School District No. 403 v. Fraser obfuscates the extent to which the First Amendment protects a college student's speech. It is arguable, especially in light of Chief Justice Burger's dissent in *Papish v. Board of Curators*[116] coupled with the conservative bent of the present Court, that college students do not enjoy the same First Amendment protections they once did.

One possibility, however, consistent with the Court's precedents rejecting a dual standard of First Amendment protection for the academic community, is that no one, not just college or high school students, enjoys the same First Amendment rights they once did. After all, Chief Justice Burger dissented in one of the Court's broadest readings of the First Amendment, *Cohen v. California*, finding the Court's interpretation of the First Amendment much too expansive.[117] It is likely the present conservative majority on the

---

110. 403 U.S. 15 (1971).

111. *See Bethel*, 478 U.S. at 682.

112. 403 U.S. 15 (1971).

113. *See Bethel*, 478 U.S. at 681.

114. *See supra* note 64 and accompanying text.

115. This duality exists despite the Court's earlier acknowledgement in *Healy* that "the precedents of this Court leave no room for the view that . . . First Amendment protections should apply with less force on college campuses than in the community at large." *Healy*, 408 U.S. at 180.

116. In *Bethel*, Chief Justice Burger essentially incorporated into the majority opinion his dissent in *Papish*. *Compare Bethel*, 478 U.S. 675 (1986) *with Papish* 410 U.S. 667, 671-73 (1973) (Burger, C.J., dissenting). The Chief Justice believed habits of civility should be taught at both the secondary and post-secondary levels of education.

117. 403 U.S. 15 (1971).

Court views the case similarly. In other words, given the right case, the Court might very well be willing to overrule *Cohen v. California*.

At present, it is clear there is a different First Amendment standard applicable to the academic community. It is, however, unclear what the exact parameters of this standard are and the extent to which it applies to post-secondary school students and teachers.

VI. RULES AND RIGHTS COLLIDING

Supreme Court precedents render some of the speech codes instituted by colleges and universities of questionable validity. Given the present makeup of the Supreme Court, however, and its tendency to limit First Amendment protections—at least at the high school level—it is arguable many speech codes would withstand Supreme Court review.[118] It may be time for society to look to colleges and universities to protect First Amendment principles instead of looking to the Supreme Court for these protections.

It seems clear that "fighting words" and advocacy that incite persons to "imminent violence" are not protected under the First Amendment and that a college or university has a legitimate interest in prohibiting such speech. Furthermore, schools may limit the time and place of speech, so long as such limitations are content-neutral. In other words, a college administration could prohibit the "student Nazis" from meeting on campus at three o'clock in the morning as long as the administration prohibited all groups from doing the same. Again, the restrictions must be content-neutral.

Additionally, a college or university, in promulgating an acceptable speech code, must ensure that the policy is not vague or overbroad.[119] Any policy limiting First Amendment rights must be particularized to insure that protected speech is not stifled and also it "must give adequate warning of the conduct which is to be prohibited and must set out explicit standards for those who apply it."[120] This is to guarantee that there is no "real and substantial chilling effect" on constitutionally protected speech.[121]

Beyond these strictures it is arguable that a college or university has very wide latitude in enacting speech codes. The Court is apparently receptive to arguments that frame the issue of limiting First Amendment rights on college campuses in terms of teaching com-

---

118. At least one federal district court has not been favorably inclined toward such codes. In *Doe v. University of Michigan*, the court found a rather detailed speech code unconstitutionally vague and overbroad. This case has not been appealed. 721 F. Supp. 852 (E.D. Mich. 1989).

119. *See id.* at 861-67.

120. Broadrick v. Oklahoma, 413 U.S. 601, 607 (1973) (citing Grayned v. City of Rockford, 408 U.S. 104, 108 (1972)).

121. *See* Young v. American Mini-Theatres, 427 U.S. 50 (1976).

municative civility. Were a public institution to show that its speech code is merely an attempt to teach students to "express themselves in acceptable, civil terms," arguably, the speech code would be considered constitutional even though such strictures might be considered unconstitutional in other settings.[122] The Court might also be receptive to arguments that would limit First Amendment speech in all settings, not just on the college campus. The present conservative Court might very well be inclined to adopt Chief Justice Burger's dissent in *Cohen v. California*.[123]

Would it not be preferable, though, for institutions to refrain from actions that might call into question their respect for First Amendment rights, regardless of those actions' constitutionality? Colleges and universities should concern themselves with engendering a campus atmosphere in which speech of all kinds flourishes and the bounds of accepted norms and principles are always tested instead of concerning themselves with the nuances of First Amendment jurisprudence. University and college administrators should be the last to restrict speech. Rather they should be the first to protect it.

Students have an interest in an unintimidating place of scholarship; however, part of scholarship is learning to cope with views that one finds abhorrent. Students' verbal battles should not be fought for them by administrators with speech codes. As one writer put it,

> The same students who insist that they be treated as adults when it comes to their sexuality, drinking and school work, beg to be treated like children when it comes to politics, speech and controversy. They whine to . . . the president or provost of the university, to "protect" them from offensive speech, instead of themselves trying to combat it in the marketplace of ideas.[124]

Students must learn tolerance for all ideas no matter how repugnant to their own beliefs. College and university administrators should not cast themselves in the role of censors.

For example, students should be able to tolerate a professor's teaching of a book from a nonfeminist viewpoint. In fact, students should take the opportunity to challenge and probe the professor as to why he or she teaches the book in such a way. It is only by verbally challenging and probing that learning is achieved. It is not through the disciplinary process and court hearings that knowledge is gained. The First Amendment cannot be restricted simply because persons are too timid to challenge ideas they disagree with or are too thin-skinned to hear any view with which they disagree.

---

122. *See Papish*, 410 U.S. at 671-72 (Burger, C.J., dissenting).

123. 403 U.S. 15 (1971). This case read very broadly the scope of the First Amendment.

124. Dershowitz, *supra* note 8, at 89.

It is only through tolerance that repugnant, false, and prejudicial ideas can truly be shown for what they are. It is through silence that such ideas gain respectability because then they are never debated and refuted. The First Amendment is not always pretty, and at times it can be rather ugly. But is it not better to expose false ideas for what they are than to have their purveyors spreading those ideas in society when they are out of school? This is not to discount the importance of teaching civility; but should not First Amendment principles come before any ill-defined notions of civility?

A student's tolerance of all speech does not include the tolerance of abhorrent actions. Certainly, there are times when the First Amendment is at odds with the Fourteenth, but when this happens persons should be punished and disciplined for their actions, not their words. Justice Brennan stated best that First Amendment rights should only be "suppressed if, and to the extent that, [speech] is so closely brigaded with illegal action as to be an inseparable part of it."[125]

For example, students that paint swastikas on fraternity houses could be punished for trespassing and defacing property. Students that chant epithets at others could be disciplined for disturbing the peace, if such disturbance results from their action. A student that harasses another should be disciplined for his or her actions, not for the words uttered.

VII. CONCLUSION

It is certainly true that the Supreme Court may be inclined to find collegiate speech codes constitutional. The existence of this tendency does not make it right. Colleges and universities, both public and private, should be fortresses where restrictions on speech are not tolerated, citadels in which the free exchange of ideas flourishes, and bastions of thought in which the limits of speech are tested.

THOMAS L. MCALLISTER

125. *See* Roth v. United States, 354 U.S. 476, 514 (1957) (Douglas, J., dissenting).

TENNESSEE'S PROHIBITION OF THE RETROACTIVE MODIFICATION OF CHILD SUPPORT ORDERS

Before 1987, Tennessee case law permitted retroactive modification of child support orders.[1] This remedy allowed trial judges to reduce or cancel support arrearages in cases where the petitioner demonstrated substantially changed circumstances, such as unemployment or disability.[2] In 1986, however, Congress enacted legislation that threatened Tennessee with the loss of Aid to Families with Dependent Children (AFDC) funds unless the state prohibited retroactive modification of support orders.[3] Tennessee complied with this federal legislation by passing Tennessee Code Annotated section 36-5-101(a)(5) in 1987.[4] The purpose of this comment is to give the reader a better understanding of Tennessee's child support law. After a general overview of the current law concerning child support and the origin of section 36-5-101(a)(5), this comment discusses the current direction of the Tennessee courts in dealing with child support arrearages.

I. INTRODUCTION

Child support debts are a serious matter. In Tennessee during 1990, more than eighty percent of all court ordered child support payments were uncollected.[5] Unlike most debts, child support debts can form the basis for criminal action resulting in imprisonment,[6] are not dischargeable in bankruptcy,[7] and may be enforced through

---

1. Dillow v. Dillow, 575 S.W.2d 289, 291 (Tenn. Ct. App. 1978); Zeitlin v. Zeitlin, 544 S.W.2d 103, 109 (Tenn. Ct. App. 1976).

2. *See Dillow*, 575 S.W.2d at 291; *Zeitlin*, 544 S.W.2d at 109; Mason v. Mason, 43 S.W.2d 1067, 1069 (1931).

3. *See* 42 U.S.C. § 666(a)(9) (Supp. 1990); *see also infra* note 32.

4. *See infra* text accompanying note 42.

5. During the federal fiscal year ending September 1991, Tennessee courts ordered $471,224,434.14 in child support payments. From this amount, only $88,995,197.06 was collected, leaving a total of $382,229,237.08 as all unpaid court ordered child support. Tennessee Department of Human Services, Child Support Services Division, OCSE-158 Report (1991).

6. TENN. CODE ANN. § 36-5-104 (1991). Section 36-5-104(a) states, "Any person, ordered to provide support and maintenance for a minor child or children, who fails to comply with the order or decree, may, in the discretion of the court, be punished by imprisonment in the county workhouse or county jail for a period not to exceed six (6) months." *Id.*

7. McMurray v. Paulson, 27 Bankr. 330, 333 (Bankr. W.D. Tenn. 1983).

the legal process of garnishment.[8] Further, reasonable attorney fees may be recovered for the enforcement of a child support decree.[9] Because child support debts carry such significant collateral disabilities, the award of child support must bear a reasonable relation to the obligor's ability to pay.[10]

In Tennessee, both parties are charged with the care, nurture, welfare, education, and support of their minor children.[11] A child has a right of support from both parents.[12] The needs of the child and the parent's ability to provide for these needs in a manner commensurate with the obligor's means and station in life are the principal criteria for the determination of child support.[13]

Because child support must be determined by what the obligor parent can pay, the federal government requires each state to adopt quantitative child support guidelines to be used in all child support decisions as the beginning point for ascertaining what is reasonable child support.[14] Guidelines apply to all Tennessee child support ordered, entered, or revised after October 12, 1989.[15] Under the Tennessee statutes, a court must determine child support payments by using the established percentage standard.[16]

---

8. TENN. CODE ANN. § 36-5-501 (1991). *But cf.* Young v. Young, 802 S.W.2d 594 (Tenn. 1990) (holding that federal supplemental security income benefits are protected from garnishment for child support).

9. TENN. CODE ANN. § 36-5-103(c) (1991).

10. *See* Plumb v. Plumb, 372 S.W.2d 771 (Tenn. Ct. App. 1963) (holding a child support award of $200 per month for support of 15-year-old daughter by divorced father whose only income was $383 per month was too high and reducing the child support award).

11. TENN. CODE ANN., § 34-1-101 (1991); Schwalb v. Schwalb, 282 S.W.2d 661, 677 (Tenn. Ct. App. 1955); Merrill v. Merrill, 216 S.W.2d 705, 706 (Tenn. 1948); Rose Funeral Home, Inc. v. Julian, 144 S.W.2d 755, 757 (Tenn. 1940).

12. TENN. CODE ANN., § 34-1-101 (1991); Brooks v. Brooks, 61 S.W.2d 654 (Tenn. 1933); Evans v. Evans, 140 S.W. 745, 746-47 (Tenn. 1911).

13. *See Plumb*, 372 S.W.2d at 776.

14. *See* 42 U.S.C. § 667 (Supp. 1991).

15. *See* TENN. CODE ANN. § 36-5-101(e)(2) (1991).

16. *See id.* The standards are quite simple to apply; one simply multiplies the obligor's net income by the applicable percentage. For example, in most cases an obligor with a support obligation for one child will pay twenty-one percent of his or her net income for child support; for two children, the obligation would be thirty-two percent of net income; for three children, forty-one percent; for four children, forty-six percent; and, for five or more children, fifty percent. *See* Walton Garrett, *Tennessee Child Support Guidelines*, 4 TENN. FAM. L. LETTER, NO. 3 JAN. 1990, AT 11, 13.

> The guidelines presume that the obligor . . . has visitation two nights on alternating weekends and two additional weeks during the year. If health care is not provided that cost should be added, plus any additional special or extraordinary expenses. To the extent that the obligor's income exceeds $6,250 per month, the excess support may be paid into an educational trust fund for the benefit of the child or children.

Id. at 13.

Generally, the percentage guidelines have two desirable effects: the amount of support provided by the non-custodial parent tends to increase, and the child support decisions in a particular jurisdiction gain predictability and uniformity.[17] Although the determination of the amount to be paid for child support is within the reasonable discretion of the court,[18] there is a rebuttable presumption in all child support cases that the amount of support determined by an application of these guidelines is the correct amount to be awarded unless the court makes a written or specific finding on the record that the application of the guidelines would be unjust or inappropriate in a particular case.[19]

II. LEGISLATIVE BACKGROUND

The current federal and Tennessee laws prohibiting retroactive modification of child support arrearages were enacted to prevent serious harm caused by non-payment of child support.[20] Governmental agencies in the United States generally have been ineffective in enforcing child support orders.[21] According to the U.S. Census Bureau, obligors comply with less than sixty percent of all child support orders.[22] Poor compliance with child support orders is a

---

17. Garrett, *supra* note 16.

18. *See* TENN. CODE ANN. § 36-5-101(e)(1) (1991).

19. *See* 42 U.S.C. § 667(b)(2) (Supp. 1991); TENN. CODE ANN. § 36-5-101(e)(1) (1991).

> Six criteria are set out for circumstances in which the guidelines would probably be inappropriate: (1) parent having income of less than $100 or more than $10,000 per month, (2) provision of significant in-kind services by a parent, (3) child having substantial income, (4) neither parent being the sole nor primary custodian, (5) unusual expenses, and (6) the Tennessee Department of Human Services having custody and the parent seeking return of custody.

Supreme Court Approves Child Support Guidelines, 4 TENN. FAM. L. LETTER, No. 5, Mar. 1990, at 1-2. If deviations occur from the guidelines, the court must consider 15 criteria for determining support:

> the child's (1) age, (2) physical condition, (3) health, (4) education, (5) educational advantages enjoyed during the parents' marriage, (6) standard of living enjoyed during the parents' marriage, (7) that which was given up, (8) income and (9) assets, the parent's (10) income and (11) assets, (12) visitation of less than 55 days or more than 110 days per year, (13) any unusual deferred compensation to be received by the parents, (14) prior child support obligations, and (15) such other factors as are necessary to consider the equities.

Id. at 2.

20. *See* Rutledge v. Barrett, 802 S.W.2d 604, 607 (Tenn. 1991).

21. Lenore J. Weitzman, *The Economic Consequences of Divorce: An Empirical Study of Property, Alimony, and Child Support Awards*, 8 Fam. L. Rep. (BNA) 4037, 4054 (Aug. 3, 1982).

22. *See id.* at 4053.

factor in what some observers call the feminization of poverty.[23] Because most custodial parents are women,[24] they bear the brunt of delinquent child support payments. Studies indicate that after a divorce a non-custodial father's standard of living rises an average of forty-two percent, while a custodial mother's standard of living drops seventy-three percent.[25] As a result, many female custodial parents and their children have been forced onto welfare roles.

Some obligors, however, are not able to comply with their support orders because of an adverse change in their economic circumstances.[26] Other obligors who are financially able to comply with their court-ordered support obligations simply refuse to pay.[27] Non-compliance for either reason shrinks the economic resources of custodial parents and their children. Hardship can result, and programs such as Aid to Families with Dependent Children (AFDC) may have to step in and make up for the money that the obligor cannot provide or refuses to pay.

Widespread non-compliance with support orders[28] has led to increased public funding of programs like AFDC,[29] stretching the resources of child support agencies and the tax-paying public. Because

---

23. For example, three out of four single mothers under the age of twenty-five live below the poverty line. Barbara D. Savage and Paula Roberts, *Unmarried Teens and Child Support Services*, 21 CLEARINGHOUSE REV. 443 (1987). Even given full compliance with child support orders, however, some custodial parents are underemployed or unemployed and may never be able to support their family (or families) with their earnings. In addition, family income that once had been adequate may become inadequate after divorce or separation because the parents no longer pool their housing, food, and transportation costs. It is thus unclear what percentage of welfare costs are due to delinquent child support payments, because chronic unemployment, discrimination and the structure of the economy may render a large proportion of welfare payments unavailable. *Id.*

24. *See* Martha L. Fineman & Anne Opie, *The Uses of Social Sciences Data in Legal Policymaking: Custody Determinations at Divorce*, 1987 WIS. L. REV. 107, 115-16. The almost automatic tendency of courts to award custody of children to the woman, however, is changing as attitudes shift and joint custody becomes more common. *Id.* at 111-21.

25. Weitzman, *supra* note 21, at 4053. Absent and unknown fathers probably have a similar effect on women's financial status.

26. Common examples of changed circumstances include unemployment, imprisonment, and disability.

27. Weitzman, *supra* note 21, at 4045.

28. *See supra* notes 5, 23, and accompanying text. The problem of child support enforcement largely stems from increases in the divorce rate, the desertion rate, and the out-of-wedlock birthrate. MICHAEL R. HENRY & VICTORIA S. SCHWARTZ, A GUIDE FOR JUDGES IN CHILD SUPPORT ENFORCEMENT 1-2 (2d ed. 1985). For example, the number of never-married mothers increased 377% between 1970 and 1983. *Id.* at 2.

29. *See* Diane Dodson and Robert M. Horowitz, *Child Support Enforcement Amendment of 1984: New Tools for Enforcement*, 10 Fam. L. Rep. (BNA) 3051 (Oct. 23, 1984).

of non-compliance with support orders, these welfare costs have drained federal and state relief funds.[30] Although enforcement of child support orders was traditionally left to the states,[31] Congress, deciding to attack the problem after ineffective state enforcement practices had resulted in a financial drain on federal AFDC funds, enacted 42 U.S.C. section 666(a)(9).[32]

III. TENNESSEE'S RESPONSE: A STATUTORY PROHIBITION OF RETROACTIVE MODIFICATION OF CHILD SUPPORT ORDERS

42 U.S.C. section 666(a)(9)[33] gave states until May 31, 1987 to enact legislation to prohibit retroactive revision of support orders and, by extension, support arrearages.[34] The federal government intended this statute to improve child support enforcement by forcing Tennessee, along with seventeen other states,[35] to enact laws like Tennessee Code Annotated section 36-5-101(a)(5).[36] Congress apparently believed that the eighteen states allowing retroactive revision of support orders created a loophole making support enforcement more difficult. Some observers suggested that prohibiting retroactive revision of support orders was desirable because it could eliminate a potential source of divisive litigation,[37] provide finality to support

---

30. Weitzman, *supra* note 21, at 4053.

31. Diann Dawson, Comment, *The Evolution of a Federal Family Law Policy Under Title IV-A of the Social Security Act—The Aid to Families with Dependent Children Program*, 36 CATH. U.L. REV. 197, 210 (1986).

32. 42 U.S.C. § 666(a)(9) provides that in order to receive federal assistance, each state must utilize procedures which guarantee that a child support payment is:

> (A) a judgment by operation of law, with the full force, effect, and attributes of a judgment of the State, including the ability to be enforced,
>
> (B) entitled as a judgment to full faith and credit in such State and in any other State, and
>
> (C) not subject to retroactive modification by such State or by any other State.

42 U.S.C. § 666(a)(9) (Supp. 1990).

33. *Id.*

34. *See States Slow to Comply with IV-D Ban on Retroactive Modification of Arrears*, 13 Fam. L. Rep. (BNA) 1270 (Apr. 7, 1987).

35. In July 1985, before the enactment of 42 U.S.C. § 666(a)(9) (1986), 11 states (Maryland, Massachusetts, Michigan, Minnesota, New Jersey, New York, Pennsylvania, Rhode Island, South Dakota, Tennessee, and Wisconsin) permitted retroactive revision of support orders by case law or statute. Seven other states (Connecticut, Delaware, Hawaii, North Carolina, South Carolina, Vermont, and Wyoming) did not expressly prohibit retroactive revision of support orders. DIANE DODSON & SHERRY GREEN DE LA GARZA, RETROACTIVE MODIFICATION OF CHILD SUPPORT ARREARS 3, A-1, A-2, & A-3 (A.B.A. National Legal Resource Center for Child Advocacy and Protection, June 1986).

36. *See infra* note 42 and accompanying text.

37. The Department of Health and Human Services had stated that allowing retroactive revision "permits arguments to be made about changed circumstances in prior periods at a time when evidence may not be abundant or clear." *HHS Rules proposed on Proscription Against Retroactive Modification of Support Arrears*, 13 Fam. L. Rep. (BNA) 1600, 1601 (Oct. 13, 1987).

orders, and encourage obligors to file modification petitions promptly.[38]

In 1987, the Tennessee Legislature enacted legislation[39] to comply with the 1986 federal amendment and thus avoid the loss of several million dollars in federal AFDC funds.[40] The legislative history of section 36-5-101(a)(5) reflects the General Assembly's "clear understanding" the courts in Tennessee would no longer be able to forgive past arrearages in child support cases but would retain their discretion to determine how and when the past due amounts were to be paid.[41] Section 36-5-101(a)(5) provides:

> Any order for child support shall be a judgment entitled to be enforced as any other judgment of a court of this state and shall be entitled to full faith and credit in this state and in any other state. Such judgment shall not be subject to modification as to any time period or any amounts due prior to the date that an action for modification is filed and notice of the action has been mailed to the last known address of the opposing parties.[42]

Before enactment of this section, Tennessee case law[43] had long permitted the obligor[44] to petition a Tennessee court for a retroactive reduction in the child support level. Such a petition, if granted, had the effect of reducing support arrearages.

The enactment of section 36-5-101(a)(5), however, overruled this long-standing Tennessee case law; the statute now permits only prospective revision of child support orders.[45] Section 36-5-101(a)(5) flatly prohibits the retroactive modification of support orders, regardless of changes in the petitioner's circumstances.[46] Instead, this statute permits a court to revise support orders only back to the date that notice of a petition to revise was given to the respondent.[47]

---

38. *See* DODSON & GREEN DE LA GARZA, *supra* note 35, at 2 (quoting Wood v. Wood, 407 A.2d 282, 287 (Me. 1979)).

39. TENN. CODE ANN. § 36-5-101(a)(5) (1991) became effective March 27, 1987.

40. Rutledge v. Barrett, 802 S.W.2d 604, 606 (Tenn. 1991).

41. *Id.* at 606 (citing Tennessee House Judiciary Committee hearing, March 3, 1987; Tennessee Senate Judiciary Committee hearing, March 3, 1987).

42. TENN. CODE ANN. § 36-5-101(a)(5) (1991).

43. Hoyle v. Wilson, 746 S.W.2d 665, 674-75 (Tenn. 1988); Dillow v. Dillow, 575 S.W.2d 289, 291 (Tenn. Ct. App. 1978); Zeitlin v. Zeitlin, 544 S.W.2d 103, 109 (Tenn. Ct. App. 1976).

44. Theoretically, the obligee could also petition for retroactive reduction of support, but such an action would be uncommon; almost all petitions to reduce support are filed by the obligor.

45. Any party may still petition for prospective revision of a support order. Such a prospective revision may either increase or decrease the support level. *See* TENN. CODE ANN. § 36-5-101(a)(5) (1991).

46. *Id.*

47. *Id.*

Because the notice date is crucial, Tennessee obligors should promptly petition for revision of their support orders as soon as their circumstances change. Any delay will result in accumulation of arrearages.

IV. CURRENT TENNESSEE LAW CONCERNING CHILD SUPPORT ARREARAGES

The first reported case in Tennessee dealing with the issue of proper construction of section 36-5-101(a)(5) was *Rutledge v. Barrett*.[48] The facts in *Rutledge* are typical of many divorce cases. The mother was awarded custody of the parties' three minor children when the parties were divorced in 1973.[49] No support order was entered against the father, who was out of state at the time of the divorce.[50] When the father returned to Tennessee in 1975, however, the mother secured a child support order requiring him to pay the court clerk one-half of his income, up to a maximum of $50.00 per week, for the support of his minor children.[51]

The father initially made a few child support payments and erratically exercised his visitation privileges until the mother left Tennessee.[52] After several years, the mother moved back to Tennessee, and again the father made a few child support payments and erratically exercised his visitation privileges.[53] Finally, in late 1985 or early 1986, the father lost all contact with the mother and children.[54] In March 1988, the mother filed a petition for contempt and asked for a judgment for the child support arrearages.[55] The Tennessee Supreme Court explored the motive and rationale behind the 1987 amendment

---

48. 802 S.W.2d 604 (Tenn. 1991). Before *Rutledge* several unreported Tennessee Court of Appeals decisions have prohibited the reduction of child support arrearages. *See* Harrison v. Smith, No. 980, 1991 WL 198901 (Tenn. Ct. App. Dec. 12, 1990); Daughtery v. Daughtery, No. 977, 1990 WL 143812 (Tenn. Ct. App. Oct. 4, 1990); France v. France, No. 56, 1990 WL 130788 (Tenn. Ct. App. Sept. 13, 1990); Podwoski v. Podwoski, No. 26, 1990 WL 88529 (Tenn. Ct. App. June 29, 1990); Friar v. Pine, No. 930, 1990 WL 74391 (Tenn. Ct. App. June 7, 1990); Thacker v. Begley, No. 270, 1990 WL 32104 (Tenn. Ct. App. Mar. 26, 1990); Caywood v. Simmons, No. 920, 1990 WL 16887 (Tenn. Ct. App. Feb. 27, 1990); Philpott v. Philpott, No. 83-306-II, 1989 WL 11871 (Tenn. Ct. App. Feb. 15, 1989).

49. *Rutledge*, 802 S.W.2d at 605. The three children were aged one, four, and eight years old at the time of the parties' divorce. *Id.*

50. *Id.*

51. *Id.* The court ordered the father to file periodic reports concerning his income to determine the proper amount of child support he was to pay. *Id.*

52. *Id.*

53. *Id.*

54. *Rutledge*, 802 S.W.2d at 605.

55. *Id.* As of March 1988, the father's child support arrearages totaled $33,555.00. *Id.* The father filed a counterclaim for contempt, claiming he had been denied visitation and asking for forgiveness of the arrearages. *Id.*

to section 36-5-101, which prohibits retroactive modification of child support orders.[56] The court stated:

> Faced with the possible loss of millions of dollars in federal assistance because of the existing retroactive modification provision in Tenn. Code Ann. § 36-5-101, the Tennessee Legislature amended that statute in 1987 to track the provisions of the new federal act. . . . The legislative history of this amendment reflects the General Assembly's clear understanding that as a result of legislative action to bring Tennessee law in line with the federal requirement, the courts of this state would lose their ability to forgive past arrearages in child support cases.[57]

After considering the rationale behind the amendment to section 36-5-101, the court held that the 1975 child support order was not subject to retroactive modification and therefore, the father's child support arrearages could not be forgiven.[58] Thus, no court in this situation could reduce the child support amounts that accrued before the filing of the father's 1988 cross-petition, including those amounts that became due before the effective date of the amendment.[59] The court also disallowed the equitable defenses[60] raised by the father as a bar to the collection of his child support arrearages.[61] According to the Tennessee Supreme Court, the only latitude courts have in dealing with child support arrearages is the discretion to determine how and when the past due amounts are to be paid.[62] Apparently, the only type of defense a respondent will have is to ask a court to allow him or her to pay minimal amounts until the child support arrearage is extinguished.[63]

The supreme court recognized that in some instances the effect of this new statute may be harsh.[64] Because of this result, the court recommended that trial courts institute appropriate procedures to provide early notice to the parties of the consequence of failing to meet obligations under a child support order.[65]

---

56. *See id.* at 606.
57. *Rutledge*, 802 S.W.2d at 606.
58. *Id.* at 607.
59. *Id.* at 606.
60. The father raised the equitable defenses of laches, estoppel, waiver, and acquiescence in trying to deny the petitioner recovery of the arrearages. *Id.* at 607. The father did not raise the statute of limitations as a defense even though the petitioner failed to bring this action until thirteen years after the father was first in arrears. *See id.*
61. *Rutledge*, 802 S.W.2d at 607.
62. *Id.* at 606.
63. *See id.* at 606-07.
64. *Id.* at 607.
65. *Id.* For example, the court stated a trial judge might inform the obligor in open court of the potentially dire consequences of the failure to pay the required

V. THE STATE OF THE LAW IN TENNESSEE AFTER *RUTLEDGE V. BARRETT*

A. Child Support Credit for Furnishing Necessaries

In *Oliver v. Oczkowicz*,[66] a Tennessee Court of Appeals unreported decision, the court noted in some instances a party may be entitled to a credit for certain payments which would offset that party's child support arrearages.[67] The parties in this case were divorced in 1984, and the mother secured custody of the two minor children.[68] According to the final decree, the respondent father was ordered to pay child support in the amount of $800.00 per month.[69] In February 1989, the mother filed a petition for child support arrearages asserting that the father owed an arrearage in the amount of $9,400.15 dating from June 1, 1984.[70] The trial court found the respondent was entitled to a credit[71] in the amount of $4,754.65 to be set off against his child support arrearage.[72]

In reversing the trial court, the *Oliver* court addressed two issues.[73] The court first discussed the "propriety of allowing a credit against accrued child support payments for voluntary expenditures on behalf

---

support. *Rutledge*, 802 S.W.2d at 607. Additionally, a court may require submission of an affidavit that informs the obligor in open court of the consequences of his failure to pay the required support. *Id.* A court may also require submission of an affidavit, prior to entry of a final judgment, indicating that the obligor has been informed by counsel of the consequences of the failure to pay the required support. *Id.*

66. No. 89-396-II, 1990 WL 64534 (Tenn. Ct. App. May 18, 1990).

67. *Id.* at *2.

68. *Id.* at *1.

69. *Id.* The father was required to pay child support in the amount of $600.00 per month for the remainder of 1984 and $800.00 per month thereafter. *Id.*

70. *Oliver*, 1990 WL 64534, at *1. The petitioner also filed a petition for current child support and medical coverage. *Id.*

The petition originally was filed in the District Court for Tarrant County, Texas pursuant to the Uniform Reciprocal Enforcement of Support Act (URESA). *Id.* The Texas court certified the claim as actionable and transferred the petition to the Tennessee Department of Human Services, which filed it in the Fourth Circuit Court for Davidson County, Tennessee. *Id.*

71. The credits in dispute were:

> (1) private school tuition in the amount of $919.50; (2) medical bills for one of the children in the amount of $100.00; (3) MasterCard charges in the amount of $1,013.50; (4) a personal loan to petitioner in the amount of $2,000.00; (5) a boy scout uniform costing $63.63; and (6) airline tickets for visitation with the children in the amount of $508.52.

Id. at *2.

72. *Oliver*, 1990 WL 64534. The court acted upon the motion by the husband for a final determination as to the credits due. *Id.*

73. *See id.* at *2.

of the children not made in accordance with the order of support."[74] Second, the court questioned the "propriety of allowing the non-custodial parent a set-off against the accrued child support payments for a debt allegedly owed by a custodial parent."[75] The court stated as a rule a party should be allowed a credit for voluntary payments made on behalf of the children only when the payments are for the children's necessaries which are not being supplied by the custodial parent.[76] Applying this rule to the facts of the case, the court held the credits given the father were not justified because they were not shown to be for the children's necessaries that the custodial parent failed or refused to furnish.[77] The trial court's action, therefore, in reducing the amount of the judgment for arrearages was a clear violation of Tennessee Code Annotated section 36-5-101(a)(5).[78]

In *Sutton v. Sutton*,[79] an unpublished Tennessee Court of Appeals decision, the court gave the father a credit that reduced his child support arrearage.[80] The parties were divorced in 1978, and the mother was awarded custody of the parties' two minor children.[81] The father was required to pay child support in the amount of $30.00 per week.[82] In February 1990, the mother filed a petition for child support arrearages asserting the father had an arrearage of $4,057.20.[83] The trial court found the father was in arrears of $4,057.20 but gave a $780.00 credit to the father because he had physical although not legal custody of one child for a one-year period.[84]

The Tennessee Court of Appeals affirmed the trial court, noting under the authority of *Oliver v. Oczkowicz*[85] they had no difficulty in concluding that the furnishing of necessaries (food, clothing and shelter) during the one-year period would more than exceed the $780.00 credit allowed.[86] The court found under the authority of *Oliver*, giving a credit for the furnishing of necessaries in this situation is not in contravention of section 36-5-101(a)(5).[87]

---

74. *Id.*
75. *Id.*
76. *Id.*
77. *Oliver*, 1990 WL 64534 at *3.
78. *Id.*
79. No. 180, 1991 WL 16234 (Tenn. Ct. App. Feb. 12, 1991).
80. *Id.*
81. *Id.*
82. *Id.*
83. *Id.* The father was constantly in arrears as to child support, and the mother was required to file numerous petitions to secure payments. *Sutton*, 1991 WL 16234.
84. *Id.*
85. No. 89-396-II, 1990 WL 64534 (Tenn. Ct. App. May 18, 1990); *see supra* notes 67-79 and accompanying text.
86. *Sutton*, 1991 WL 16234.
87. *See id.* at *1.

In another Tennessee Court of Appeals unpublished decision, *Dalton v. Dalton*,[88] the court gave a credit to the father for child support payments that lacked compliance with the divorce decree.[89] According to the parties' divorce decree, the father was required to pay $375.00 per month child support to the mother.[90] The father had an arrearage totaling $3,140.00.[91] The chancellor found that one of the children had been living with his grandmother and the father had been paying the grandmother for the support of that child.[92] The chancellor gave credit to the father in the amount of $1,140.00 for support paid on behalf of the child to the grandmother.[93]

The Tennessee Court of Appeals affirmed the chancellor, holding that under the authority of *Oliver*[94] a credit should be allowed to the father against his child support arrearage, and such a credit is not in violation of section 36-5-101(a)(5).[95] The court noted that although the divorce decree contemplated that the children would reside with the mother, the proof was undisputed that for a significant period of time and with the consent of the custodial parent, one child lived with its grandmother.[96] Because the proof was also undisputed that the father paid child support directly to the grandmother for the support of the child, the father should be given credit for those child support payments.[97]

B. Statute of Limitations Defense

Tennessee Code Annotated section 28-3-110(2) provides "[a]ctions on judgments and decre[e]s of courts of record" must be commenced within ten years.[98] The question arises whether this ten-year statute of limitations could bar a claim for the collection of child support arrearages.[99]

---

88. No. 28, 1991 WL 25924 (Tenn. Ct. App. Mar. 1, 1991).

89. *Id.* at *1.

90. *Id.* The amount of child support was to be prorated at one-third per child. *Id.*

91. *Id.*

92. *Dalton*, 1991 WL 25924 at *1.

93. *Id.* The Chancellor entered judgment in favor of the mother in the amount of $2,000.00 for the father's arrearage. *Id.* The father was required to pay $50.00 per month to the mother on the arrearage. *Id.*

94. No. 89-396-II, 1990 WL 64534 (Tenn. Ct. App. May 18, 1990); *see supra* notes 66-78 and accompanying text.

95. *Dalton*, 1991 WL 25924 at *2.

96. *Id.* at *1.

97. *See id.*

98. TENN. CODE ANN. § 28-3-110(2) (1980).

99. *See* Sandidge v. Brown, No. 03A01-9104CV142, 1991 WL 167149 (Tenn. Ct. App. Sept. 3, 1991) (child support payments are excluded from the ten-year statute of limitations); Pera v. Peterson, No. 72, 1990 WL 200582 (Tenn. Ct. App.

The effect of this statute of limitations on a claim for child support arrearages is analyzed in *Sandidge v. Brown*,[100] an unpublished Tennessee Court of Appeals decision. In December 1975, the putative father of a minor child was ordered to pay $30.00 per week child support.[101] In June 1990, the mother filed an affidavit asserting the father was in arrears on child support payments totaling $21,530.00.[102] Without denying the arrearage, the father insisted the amount that had accrued more than ten years before the mother's August 1990 filing of the affidavit was barred by the ten-year statute of limitations.[103] The Court of Appeals affirmed the trial court's finding that the ten-year statute of limitations codified in Tennessee Code Annotated section 28-3-110(2) was not applicable to child support arrearages of minor children.[104]

The Tennessee Court of Appeals noted that under the authority of *Rutledge v. Barrett*[105] the obligation of support for minor children is continuing, and the child's right to support cannot be defeated by the running of the statute of limitations because of the custodial parent's dereliction in enforcing a support decree.[106] According to the court, the legislative history of section 36-5-101(a)(5)[107] shows that the Tennessee General Assembly, in adopting this amendment, intended to enhance the effectiveness of child support orders, and that the amendment was written to protect the child's interest in the support payments, rather than to create possible defenses for the

---

Dec. 14, 1990) (the obligation of support for minor children is continuing, and the right of the child to support cannot be defeated by the running of the statute of limitations due to the dereliction of the custodial parent in enforcing a decree of support); State *ex rel* Woody v. Morris, No. 44, 1990 WL 2867 (Tenn. Ct. App. Jan. 19, 1990) (ten-year statute of limitations does not bar a claim brought by the state for child support payments); Vaughn v. Vaughn, No. 88-26-II, 1988 WL 68062 (Tenn. Ct. App. July 1, 1988) (the child's right to support is not defeated upon expiration of the statute of limitations due to failure of the custodial parent to pursue a support decree); *cf.* Roberts v. Roberts, No. 1355, 1990 WL 130816 (Tenn. Ct. App. Sept. 12, 1990) (court entered judgment for all child support payments not barred by the ten-year statute of limitations).

100. No. 03A01-9104CV142, 1991 WL 167149 (Tenn. Ct. App. Sept. 3, 1991).

101. *Id.* at *2.

102. *Id.*

103. *Id.*; *see* TENN. CODE ANN. § 28-3-110(2) (1980); *see also supra* text accompanying note 98. The father argued that the arrearage between 1975 and 1980 was barred by the statute, and therefore his arrearage totaled $15,600.00 instead of $21,530.00. *Sandidge*, 1991 WL 167149 at *1.

104. *Sandidge*, 1991 WL 167149 at *3.

105. 802 S.W.2d 604 (Tenn. 1991); *see supra* notes 48-65 and accompanying text.

106. *Sandidge*, 1991 WL 167149 at *2; *see* Pera v. Peterson, No. 72, 1990 WL 200582 (Tenn. Ct. App. Dec. 14, 1990); Woody v. Morris, No. 44, 1990 WL 2867 (Tenn. Ct. App. Jan. 19, 1990); Vaughn v. Vaughn, No. 88-26-II, 1988 WL 68062 (Tenn. Ct. App. July 1, 1988).

107. *See supra* notes 41-42 and accompanying text.

obligor.[108] The court held that the ten-year statute of limitations codified in section 28-3-110(2) does not bar an action for child support arrearages.[109]

In *Roberts v. Roberts*,[110] an unpublished Tennessee Court of Appeals decision holding opposite *Sandidge*, the court barred all child support arrearages not falling within the state's ten-year statute of limitations.[111] In *Roberts*, the parties were divorced in 1971, and exclusive custody of the two minor children was awarded to the mother.[112] The father was ordered to pay $30.00 per week for partial support of the children.[113] The father paid no support and had no contact with his former wife or his children.[114] The mother filed a petition in March 1989 seeking recovery of the delinquent support payments.[115] The trial judge entered a judgment for all support payments not barred by the ten-year statute of limitations.[116]

The Tennessee Court of Appeals affirmed the trial court.[117] The court held the 1971 judgment for child support was subject to the ten-year statute of limitations pursuant to section 28-3-110.[118] The court also noted that the forgiveness of arrearages was now impermissible and that under section 36-5-101(a)(5),[119] virtually all equitable defenses to eliminate child support arrearages are precluded.[120] The court held, however, that a child support judgment remains subject to the ten-year statute of limitations.[121]

VI. CONCLUSION

The preceding material attempts to outline Tennessee child support law, with an emphasis on the statutory prohibition of retroactive modification of support orders.[122] While this statute was enacted to

---

108. *Sandidge*, 1991 WL 167149 at *2.
109. *Id.* at *3.
110. No. 1344, 1990 WL 130816 (Tenn. Ct. App. Sept. 12, 1990).
111. *Id.* at *1.
112. *Id.*
113. *Id.*
114. *Id.*
115. *Roberts*, 1990 WL 130816. In December 1988, the children filed a petition for contempt against their father to recover the delinquent support payments, and this petition was later "amended to include their mother, in effect, as a petitioner." *Id.* at *1 & n.1. The mother's petition, filed in March 1989, was adjudicated by the trial court. *Id.* at *1.
116. *Id.*
117. *Id.* at *2.
118. *Roberts*, 1990 WL 130816 at *2.
119. *See supra* note 42 and accompanying text.
120. *Id.*
121. *Id.*
122. *See* TENN. CODE ANN. § 36-5-101(a)(5) (1991); *see also supra* text accompanying note 42.

prevent serious harm resulting from non-payment of child support, it will also cause harm to obligors who have genuinely experienced a change of circumstances and who lack adequate access to the courts. Although section 36-5-101(a)(5) may have the effect of encouraging timely petitions to revise, it could lead to harsh results if the obligor's circumstances have changed long before notice is given. Obligors with inadequate access to attorneys and the legal system, such as institutionalized persons, the poor, and the unemployed, are likely to be both unaware of section 36-5-101(a)(5) and unable to give prompt notice. Consequently, such obligors will accumulate support arrearages that are extremely difficult to expunge, and which may have resulted from their changed circumstances and inadequate access to the courts rather than from their willful refusal to pay support.

One way to offset this problem is to allow immediate income withholding on all new or revised support orders. Under this collection system, employers immediately withhold an obligor's child support payment from the obligor's pay check. The employer then sends the payment directly to the clerk of the courts, who registers the payment.[123]

This system eliminates the accumulation of arrearages (unless the obligor is concealing income[124]), particularly when expressed as a percentage of the obligor's income in the court order. The child support guidelines could be used to adjust for changes in the obligor's financial condition. To the extent such an automatic adjustment of support occurs, the guidelines and immediate income withholding would eliminate the need for retroactive expungment of support arrearages.

Alternatively, the problem of obligors lacking adequate access to legal assistance and the courts might be addressed by allowing parties to enter into settlement agreements to soften the effect of section 36-5-101(a)(5) for those obligors with good cause for retroactive modification of a child support order. In a settlement agreement, the obligee must be willing to stipulate that payment of a certain sum will satisfy the debt represented by the arrearage. If the judge approves the settlement agreement,[125] the obligor's debt is satisfied

---

123. This arrangement would serve two practical functions. First, support money reaches the custodial household promptly. Second, the system provides the clerk of courts with evidence of all support payments and thus eliminates disputes about whether the obligor has paid the ordered support money.

124. Of course, immediate income withholding works only if income is reported; obligors who do not report money they earn and otherwise conceal income will not have any of their unreported income withheld for child support.

125. January 1, 1992, TENN. CODE ANN. § 36-5-101(h) (1991) was amended as follows:

In any such agreement, the parties must affirmatively acknowledge that no

upon payment of the agreed amount of money to the obligee.

Section 36-5-101(a)(5) does not explicitly forbid courts from approving such settlement agreements between obligee and obligors. A plausible interpretation of section 36-5-101(a)(5) is that the statute is merely a limitation on the court's ability to order retroactive modification. Under section 36-5-101(a)(5), courts remain free to approve settlement agreements between parties.

TIMOTHY M. MCLAUGHLIN

---

action by the parties will be effective to reduce child support after the due date of each payment, and that they understand that court approval must be obtained before child support can be reduced, unless such payments are automatically reduced or terminated under the terms of the agreement.

TENN. CODE ANN. § 36-5-101(h) (1991).

RECONSTRUCTING LIBERTY

ROBIN WEST*

INTRODUCTION

It is commonly and rightly understood in this country that our constitutional system ensures, or seeks to ensure, that individuals are accorded the greatest degree of personal, political, social, and economic liberty possible, consistent with a like amount of liberty given to others, the duty and right of the community to establish the conditions for a moral and secure collective life, and the responsibility of the state to provide for the common defense of the community against outside aggression. Our distinctive cultural and constitutional commitment to individual liberty places very real restraints on what our elected representatives can do, even when they are acting in what all of us, or most of us, would consider our collective best interest. For example, we cannot outlaw marches by the Ku Klux Klan,[1] or the burning of flags by political extremists,[2] or the anti-Semitic, racist, or hateful speech of incendiary and potentially dangerous bigoted zealots.[3] Nor can we simply outlaw those practices of religious sects that may have deleterious effects on the members, such as the refusal of certain Amish sects in the Eastern United States to allow

---

* Professor of Law, Georgetown University Law Center. B.A., 1976, University of Maryland; J.D., 1979, University of Maryland; J.S.D., 1982, Stanford Law School. This article was originally a speech delivered on September 6, 1991, as the Alumni Distinguished Lecture in Jurisprudence at the University of Tennessee College of Law in Knoxville, Tennessee. Early drafts of this speech were presented at Notre Dame College in Baltimore, Maryland; The American Philosophical Association Meeting in New York, New York, January 1992; George Washington Law School; and University of Maryland, College Park, Maryland. I thank the participants in those colloquia for their comments and criticisms.

1. *See* Collin v. Smith, 578 F.2d 1197 (7th Cir.) (ordinances in Skokie, Illinois, designed to block march by Ku Klux Klan found unconstitutional), *cert. denied*, 439 U.S. 916 (1978).

2. *See* United States v. Eichman, 496 U.S. 310 (1990) (federal statute criminalizing flag desecration found unconstitutional); Texas v. Johnson, 491 U.S. 397 (1989) (same result).

3. *See In re R.A.V.*, 464 N.W.2d 507 (Minn.), *cert. granted*, 111 S. Ct. 2795 (1991). For arguments to the effect that hate speech should not be constitutionally protected, see Richard Delgado, *Words that Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling*, 17 HARV. C.R.-C.L. L. REV. 133 (1982); Charles R. Lawrence, III, *If He Hollers Let Him Go: Regulating Racist Speech on Campus*, 1990 DUKE L.J. 431; Mari J. Matsuda, *Public Response to Racist Speech: Considering the Victim's Story*, 87 MICH. L. REV. 2320 (1989). *But see* Nadine Strossen, *Regulating Racist Speech on Campus: A Modest Proposal?* 1990 DUKE L.J. 484.

their children to receive a public education past the eighth grade,[4] the explicit exclusion (until recently) of blacks from positions of influence in the Mormon Church, or the continuing exclusion of women from positions of power, prestige, and influence in our dominant, mainstream, Protestant, Catholic, and Judaic faiths. We may believe correctly that a full civic education for every individual is not only desirable for its own sake but is an absolute prerequisite for meaningful participation in our shared political life. We may believe that racist speech is antithetical to the racial tolerance necessary to our continued existence as a pluralistic society, that flag-burning communicates no message worth hearing, and that women and blacks are entitled to the opportunity to aspire to positions of full participation and responsibility in religious life. Nevertheless, we are precluded from legislating in a way that would put the weight of the law behind these values because to do so ostensibly would do great violence to something we hold even more dear: the right and responsibility of the individual to think, speak, and act autonomously in matters of religious, political, and social life—to reach one's convictions on one's own and for oneself, unfettered by the moral dictates of the state, even where those dictates are benign and wise.

In constitutional discourse, this complex aspiration is often captured by the phrase "ordered liberty."[5] The first thing to note about this aspiration of ordered liberty is that it is a relatively modern and distinctively liberal interpretation of our constitutional heritage. Thus, although Justice Cardozo coined the phrase "ordered liberty" in the 1930s,[6] our modern understanding of ordered liberty protected by the Constitution came to full fruition with the liberty-expanding cases of the liberal Warren Court era. Quite possibly it received its most definitive formulation in the 1960s case *Poe v. Ullman*.[7] Dissenting in *Poe*, Justice Harlan wrote:

---

4. *See* Wisconsin v. Yoder, 406 U.S. 205 (1972). In *Yoder*, after balancing the state's interest in education against impingement on fundamental right of Amish to raise children in manner consistent with religious precepts, the Court held unconstitutional a parent's conviction for refusing to send a child to a public school past the eighth grade. *Id.* at 234.

The Rehnquist Court, however, may be moving away from the general principle cited in the text. *See* Employment Div. v. Smith, 494 U.S. 872 (1990) (Oregon statute criminalizing nonrecreational drug use held not to infringe First Amendment rights of Native American Church members absent showing of specific intent to burden the minority religion).

5. The phrase apparently originated in Justice Cardozo's majority opinion Palko v. Connecticut, 302 U.S. 319 (1937). In *Palko*, the Court held that the kind of double jeopardy risked by a state statute permitting the state to appeal criminal cases (1) did not "violate those 'fundamental principles of liberty and justice which lie at the base of all our civil and political institutions,'" *id.* at 328 (citing Hebert v. Louisiana, 272 U.S. 312, 316 (1926)); (2) was not "of the very essence of a scheme of ordered liberty," *id.* at 325; and (3) was not unconstitutional. *Id.*

6. *Id.* at 325.

7. Poe v. Ullman, 367 U.S. 497 (1961).

> [Implicit in the concept of ordered liberty are] those rights "which are . . . *fundamental*; which belong . . . to the citizens of all free governments" . . . for "the purposes [of securing] which men enter into society" . . . .
>
> Due process [which protects such ordered liberty] has not been reduced to any formula; its content cannot be determined by reference to any code. The best that can be said is that through the course of this Court's decisions it has represented the balance which our Nation, built upon postulates of respect for the liberty of the individual, has struck between that liberty and the demands of organized society. . . .
>
> . . . [T]he liberty guaranteed by the Due Process Clause . . . is not a series of isolated points [represented by the Bill of Rights] . . . . It is a rational continuum which, broadly speaking, includes a freedom from all substantial arbitrary impositions and purposeless restraints . . . and which also recognizes, what a reasonable and sensitive judgment must, that certain interests require particularly careful scrutiny of the state needs asserted to justify their abridgment.[8]

To paraphrase a bit, our modern understanding of ordered liberty implies that the state may not interfere with the personal or individual decisions that are most fundamental to a free life or with those liberties the protection of which is what prompts individuals—or would prompt individuals if given the explicit option—to enter civic society in the first place. The driving idea behind this notion of ordered liberty is that the protection of those liberties by the state against its own tendency to intrude in the name of some shared political end is of a higher order or of greater importance to civic life than any other conceivable and temporal state goal. Which particular liberties we view as fundamental and hence requiring this constitutional protection against even wise and benign state regulation is, of course, a subject of deep and profound disagreement. There is, however, a remarkably broad consensus in our contemporary legal culture and in our national community generally about the quite modern and quite liberal idea or aspiration of ordered liberty: that there are some liberties, whatever they may be, so essential to an autonomous life that they must be kept free of state control.

In my comments, I will be largely critical of this understanding of ordered liberty, which I occasionally will call the "modern" or "liberal" interpretation of our constitutional heritage. I want to make two objections to this concept of liberty, one political and one historical. The political objection is that the modern conception of ordered liberty is a largely empty promise for women. My claim, very briefly, will be that even the ideal expressed by this conception of ordered liberty—to say nothing of the actual practices it protects—

---

8. *Poe*, 367 U.S. at 541-43 (Harlan, J., dissenting) (citations omitted).

is skewed against women in a significant manner. The historical objection is that the liberal conception of liberty is also a cramped, inaccurate understanding of our constitutional history. I will conclude by arguing that we could fundamentally reconceive liberty in a more generous and explicitly feminist way without doing violence to either liberalism or to the document we have inherited.

Before I embark on the main project, however, one preliminary comment is in order. I want to emphasize at the outset what I am not doing. By embracing a critical posture toward the generally liberal concept of ordered liberty so eloquently spelled out by Justice Harlan above and by advocating in its stead a quite different conception, I am not endorsing, and fervently hope not to be understood as endorsing, the conservative critique of ordered liberty presently being urged in a number of opinions by Justice Antonin Scalia of the United States Supreme Court.[9] Furthermore, I do not mean to embrace the very different conception of that ideal being developed in a disturbingly large and growing number of recent Supreme Court decisions.[10] My general aim is to argue that the liberal understanding of ordered liberty articulated by Justices Cardozo and Harlan and given full meaning by the liberty-expanding cases of the Warren Court era is unduly cramped and ungenerous. It does not go far enough to do what it purports to do on its own terms, which is to protect the autonomy and liberty of individuals. Specifically, it does not protect the autonomy and liberty of women.

Justice Scalia's critique is quite the opposite.[11] Justice Scalia, and to a lesser extent his fellow conservative colleagues on the Court,

---

9. *See* Pacific Mut. Life Ins. Co. v. Haslip, 111 S. Ct. 1032 (1991); Michael H. v. Gerald D., 491 U.S. 110 (1989). In both cases, Justice Scalia argues that the liberty protected by substantive due process should be limited to those liberties historically and traditionally protected against precipitous majoritarian abridgment.

10. *See, e.g.*, Rust v. Sullivan, 111 S. Ct. 1759 (1991) (regulation forbidding federally funded clinics to counsel regarding abortion does not violate constitutional right of privacy or free speech); Hodgson v. Minnesota, 110 S. Ct. 2926 (1990) (state statute requiring notice to both parents regarding abortion request by minors held constitutional if accompanied by judicial bypass); Employment Div. v. Smith, 494 U.S. 872 (1990) (state statute criminalizing nonrecreational drug use does not violate First Amendment rights of Native American Church members); Webster v. Reproductive Health Servs., 492 U.S. 490 (1989) (state statutes regulating abortion held constitutional; *Roe v. Wade* trimester scheme explicitly questioned and arguably overruled); Bowers v. Hardwick, 478 U.S. 186 (1986) (statute criminalizing homosexual or heterosexual sodomy does not violate constitutional norms of privacy).

11. According to Justice Scalia, the most important things the Court should protect in the name of the liberty protected by the due process clause of the Fourteenth Amendment are not the private decisions that occur in the spheres of life necessary to the preservation of true individual autonomy, but rather, the decisions or spheres of life that historically and traditionally have been understood as insulated against state encroachment. *See Haslip*, 111 S. Ct. at 1032; *Michael H.*, 491 U.S. at 100. It should be apparent at once that this is a far narrower

clearly believes that the liberal understanding of ordered liberty as tied to the fundamental needs and interests of the ideally autonomous individual is too generous toward the individual. He believes that it has unduly limited the sphere of legitimate state control of individual liberty and privacy and has granted the individual too much freedom vis-à-vis the community and state within which the individual must live and be a part. Accordingly, Justice Scalia and his conservative colleagues want to shrink the sphere of ordered liberty so as to guarantee less liberty and provide more order. By contrast, I would like to see us expand that sphere. Unlike Justice Scalia's attack, my critique of the liberal understanding of ordered liberty is decidedly friendly.

As the center of power on the Court shifts from the liberal bloc of the Warren-Burger years to the conservative bloc of the Rehnquist-Scalia years, it becomes less clear, of course, what role friendly critiques such as the one I intend to offer are to play in our constitutional conversations. We are at this moment occupying an

---

concept of ordered liberty than that articulated by Justice Harlan in *Griswold*. *See Griswold*, 381 U.S. at 501-02. Instead, it implies a very different and much more limited conception of what, concretely, must be protected against state encroachment. Stating the idea in the negative, under Justice Scalia's analysis unless a sphere of decisionmaking has been historically and traditionally protected, it is not a part of the liberty protected against state encroachment. For example, according to this approach, neither the so-called liberty to engage in extra-marital sex, premarital sex, same-sex relations, or nonreproductive sex, *cf. Bowers*, 478 U.S. at 192-95, nor the liberty to procure a legal abortion, *see Rust*, 111 S. Ct. at 1776-78; *Hodgson*, 110 S. Ct. at 2961-72; *Webster*, 492 U.S. at 532-37, nor the liberty of a father to pursue a relationship with a child born to a woman married to another man, *see Michael H.*, 491 U.S. at 118-30 (Justice Scalia writing for the Court), nor the liberty of worshippers in the Native American Church to ingest peyote as part of religious rituals, *see Smith*, 494 U.S. at 876-90, are a part of the liberty constitutionally protected against state encroachment (although all such liberties would be protected under Justice Harlan's account of liberty). This is because historically and traditionally we have not protected these decisions, regardless of whether or not sexual life, parental responsibility, or spiritual practices are spheres of decisionmaking central to individual autonomy. That we have not historically and traditionally protected these liberties, of course, is evidenced by the existence of the sodomy laws, fornication laws, prohibitions against homosexuality, and criminalization of abortion and nonrecreational drug use challenged in these and similar cases. The decisions from the Rehnquist Court over the last five years, partially embracing Justice Scalia's approach and truncating or abolishing a wide range of individual liberties, substitute tradition for the liberal understanding of autonomy as the criterion for determining whether an individual liberty must be protected. This marks a profound turning point in the development of our conceptual understanding of what ordered liberty requires.

I discuss the difference between Justice Scalia's approach to liberty and that of Justice Brennan and the Warren Court generally in Robin L. West, *The Ideal of Liberty: A Comment on* Micheal H. v. Gerald D., 139 U. PA. L. REV. 1373 (1991). *See also Michael H.*, 491 U.S. at 136-56 (Brennan, J., dissenting) (defending the more liberal Warren Court approach).

ambiguous historical moment with regard to very basic constitutional norms. It is not clear whether the liberal understanding of ordered liberty briefly spelled out above will survive the conservative revolution on the Court presently underway. Should Justice Scalia's reformulation of ordered liberty—according to which the Constitution protects, in the name of liberty, the traditions of our collective past rather than the decisions of an ideally autonomous and individual life—prove successful, then so-called friendly critiques of the liberal understanding of ordered liberty may become, in a constitutional sense, simply beside the point.

On the other hand, if the liberal understanding of ordered liberty does survive, then it is imperative that we criticize it and try to improve upon it. What I am calling the liberal conception of ordered liberty does still dominate constitutional discussion, interpretation, and doctrine. It is still the ruling doctrine and is still a fundamental part of our constitutional law. It should go without saying (although in this time of hyper-patriotism it unfortunately often does not) that we best honor the Constitution and the law we create under it not by blindly revering its doctrines and certainly not by pledging our loyalty to its present form, but by interpreting it, struggling with it, criticizing it, setting its goals against itself, and forcing it and us to be true to our noblest selves. To the extent that the concept of ordered liberty elaborated by the liberal Court during the Warren Court years is still a part of the law that governs us, we should subject it to criticism so as to improve upon it, the principles it articulates, and the societal practices it governs.

Even if the conservative Court succeeds in replacing the liberal aspiration of ordered liberty honored by the Warren Court with the very different set of conservative aspirations urged by Justice Scalia, friendly critiques of the liberal concept of ordered liberty are still important to make and hear. The aspiration of ordered liberty imperfectly implemented by the great liberal decisions of the Warren Court is not only a constitutional aspiration, important as constitutionalism may be, but also a cornerstone of modern liberal theory. As a part of the political theory and of the utopian dream we call liberalism, a dream that predates and heavily informs our constitutional ideas and practices, it behooves us to "get it right." We should strive to make our conception of ordered liberty the best it can be, even if the liberalism of which it is a part survives as only a dissident voice, rather than a living part, of our positive constitutional law.

I. ORDERED LIBERTY

The liberal and relatively modern conception of ordered liberty I want to address has at least two salient features. First, the regime of ordered liberty to which we aspire is, to use Isaiah Berlin's famous

formulation, a regime of negative rather than positive liberty.[12] It is liberty or freedom from, not liberty or freedom to, which the Bill of Rights protects. When we speak of ordered liberty, we speak of the individual's liberty or freedom from invasion, intrusion, intermeddling, or over-regulation rather than the positive liberty or freedom to live a particular way, to attain one's full potential, actualize one's inner nature, or even govern oneself in a well-run democratic or majoritarian system.[13] We generally are not concerned, in our constitutional aspiration to ordered liberty, with the freedom that comes from being well-fed, clothed, sheltered, educated, or actively participating in the laws that govern us.[14] We are concerned instead with the freedom to be ourselves within some defined sphere—the freedom to make our own decisions, think our own thoughts, worship our own deities, and choose our own way of life within some sphere the boundaries of which admittedly are not clearly discernible but which are absolutely inviolable once drawn. We are concerned with the right to be left alone[15] and not with the right to any particular way to be. Where those boundaries within which we have the right to be left alone are to be drawn will be and must be a function of our known human nature and, as such, will be debated endlessly. That the boundaries must be drawn somewhere, however, is the very essence of the liberal interpretation of our Constitution as well as, perhaps, the very essence of modern liberalism. The political philosopher Isaiah Berlin describes negative liberty in this way:

> [S]ome portion of human existence must remain independent of the sphere of social control. To invade that preserve, however small, would be despotism. . . . We must preserve a minimum area of personal freedom if we are not to "degrade or deny our nature". . . . What then must the minimum be? That which a man cannot give up without offending against the essence of his human nature. What is this essence? What are the standards which it entails? This has been, and perhaps always will be, a matter of infinite debate. But whatever the principle in terms of which the area of non-interference is to be drawn, whether it is that of natural

---

12. *See* SIR ISAIAH BERLIN, *Two Concepts of Liberty, in* FOUR ESSAYS ON LIBERTY 118 (1969).

13. *Id.* at 121-22.

14. Thus, so-called welfare rights are not protected constitutionally. *See, e.g.*, Harris v. McRae, 448 U.S. 297 (1980) (Hyde Amendment prohibiting federal funding of abortions for the poor upheld as constitutional); San Antonio Sch. Dist. v. Rodriguez, 411 U.S. 1 (1973) (no constitutional right to an education). *See generally* Frank I. Michelman, *Welfare Rights in a Constitutional Democracy*, 1979 WASH. U. L.Q. 659 (1979) [hereinafter *Welfare Rights*]; Frank I. Michelman, *Foreword: On Protecting the Poor Through the Fourteenth Amendment*, 83 HARV. L. REV. 7 (1969).

15. Samuel Warren and Louis Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).

> law or natural rights, or of utility or the pronouncement of a categorical imperative, or the sanctity of the social contract, . . . liberty in this sense means liberty *from*; absence of interference beyond the shifting, but always recognizable, frontier.[16]

With only a few exceptions, most notably the right to vote guaranteed by the Fifteenth Amendment, the ordered liberty that the Constitution protects, according to the modern conception, is our negative liberty to be left alone and not our positive liberty to food, shelter, a job or an income, or to a fulfilled, prosperous, meaningful, and self-governed life. The constitutional preference for negative over positive liberty is captured by the oft-made claim that the Constitution itself is a negative one. The Constitution, it is said, protects our negative rights to be free from intrusion instead of our positive rights to a positively free, active, involved, civic, or healthy existence. The Constitution at least according to its modern interpreters, is a shield of protection; it is not a sword of entitlement.

The second feature of the modern conception of ordered liberty has its origins not in liberal theory, but in constitutional doctrine. By ordered liberty, we aspire to a regime that respects the negative freedom of the individual, and more specifically, to a regime that respects the negative freedom of the individual from undue intermeddling or interference from one and only one source—the state.[17]

---

16. BERLIN, *supra* note 12, at 126-27. Although it is a common belief that negative liberty and positive liberty are two sides of the same coin or in some way are correlated with each other, this need not be the case, as Berlin tried to show in his famous essay. *Id.* at 131. A society can be rich in one kind of liberty but poor in the other. For example, as individual citizens, we might enjoy a great deal of negative freedom such as the right to speak, worship, or be free of arbitrary arrest even though we live in a virtual dictatorship. A dictator may decide in the interest of stability or for relatively more benign reasons to grant citizens a broad sphere of inviolable freedom within which they may do as they please, even though they have no say in the governance of the society, no vote, and no right to political representation or participation. In such a society, the individual would enjoy extensive negative liberty but no positive liberty.

On the other hand, a society might be a perfectly functioning democracy, in fact as well as theory, yet it may grant absolutely no negative freedom to the individual citizen. This was the possibility that major classical liberal thinkers from Mill to Berlin both saw and feared in western democracies. A governing majority, perfectly representative of the public's will, might decide to strip individuals of all negative freedom and dictate on ideological grounds what individuals should think and believe, what they should read, and how and who they should worship. Such a society might be rich in positive freedom but poor in negative freedom. As Mill insisted, insuring to each and every individual an equal power to oppress others is no guarantee of liberty. A majority, no less than a tyrant, can squelch the negative freedom necessary for individuality, genius, creativity, spontaneity, and life itself to flourish. *See generally* JOHN STUART MILL, ON LIBERTY (1859).

17. *See, e.g.*, DeShaney v. Winnebago County Dep't of Social Servs., 489 U.S. 189 (1989) (constitutional guarantee to liberty triggered by state action, not by mere inaction); Flagg Bros. v. Brooks, 436 U.S. 149 (1978) (same).

Accordingly, the Constitution, the document on which we rely to give teeth to our aspirations, overwhelmingly is concerned with the potential for oppression in the relationship between the individual and both federal and state government. It does not, then, address the potential for oppression between the individual and other forms of organized social authority, such as the corporate employer, the trade union, the family, or the church, which also may infringe upon an individual's negative freedom. In other words, the constitutional dictate of ordered liberty places limits only on the state's potential for control. This second principle is what is often referred to in constitutional doctrine as the state action requirement. The constitutional guarantee of negative liberty is not triggered unless the state has acted in some way that infringes a protected and fundamental right. If we put these two principles together the modern conception of ordered liberty means that the Constitution protects the negative liberty of the individual against excessive intrusion by the state, by state officials, or, at the outer extreme, authorities acting under color of state authority.

Constitutional law is an admittedly complex subject, and the following generalities are subject to a host of exceptions. Nevertheless, from these two basic premises—that the liberty protected by our constitutional aspirations is negative, rather than positive, and that it is only liberty from state action and not liberty from other sources of social authority that is protected—we can generate not only much of the modern content, but more importantly for these purposes, most of the limits of our specific constitutional guarantees. From the first principle—that the Constitution protects negative rather that positive liberty—we can generate the limits the Court has imposed on the *substance* of the rights that the Constitution protects. We are guaranteed the freedom to speak, believe, associate or not associate with others, but we are not guaranteed an education,[18] adequate shelter, clothing, food, a job, or an income.[19] The former are negative freedoms while the latter, often called welfare rights, are examples of positive liberties and, hence, not protected.[20] We are guaranteed the freedom to send our children to a private school of our choice, if we can afford it, free of state interference to the contrary because this is easily characterized as a negative freedom.[21]

---

18. San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 35 (1973).

19. On the need to create constitutional entitlements to these so-called welfare rights and arguments to the effect that the Constitution should guarantee them, see PATRICIA J. WILLIAMS, THE ALCHEMY OF RACE AND RIGHTS: THE DIARY OF A LAW PROFESSOR (1991); Michelman, *Welfare Rights*, *supra* note 14.

20. *See generally* Michelman, *Welfare Rights*, *supra* note 14, at 659-60.

21. *See* Pierce v. Society of Sisters, 268 U.S. 510, 535 (1925) (finding statute requiring public rather than private education of children an unconstitutional infringement of the "liberty of parents and guardians to direct the upbringing and education of children under their control").

It is a part of the very general freedom to contract as one pleases without state interference—perhaps the quintessential negative liberty in a market economy—as well as the right to raise one's children according to one's own preferences without state interference.[22] We are not, however, guaranteed the right to a private school education regardless of ability to pay or even to a quality public school education.[23] The so-called right to an education is a positive freedom and, therefore, is not protected. We are still guaranteed (albeit narrowly) the right to procure contractually an abortion,[24] if we can afford one, because this is a negative freedom and part of our right to be left alone. We are not guaranteed the right to an abortion whether or not we can pay for it because that would be a positive freedom and would not be protected.[25] We are not even guaranteed the right to abortion counseling, for that, too, would be a positive right and, hence, not protected.[26] We are (more or less) guaranteed the right to read whatever we wish within the confines of our own home, but we are not guaranteed the right to literacy. The former is part of the negative right to be left alone while the latter is, if anything, part of a positive concept of liberty. The general rule I am suggesting is this: The Constitution guarantees us the right to do certain things free of interference from social authority, but it does not guarantee us the absolute right to do those same things. The negative freedom that is the concern of the Constitution extends only to the right to procure goods or develop abilities free of interference from social authority. It does not positively guarantee the individual the goods themselves or access to the goods or access to the ability or skills necessary to procure them.

We can generate the limits of the *scope* of the rights the Constitution protects from the second principle—that the negative freedom which is the concern of the Constitution extends only to negative freedom against interference from the state, what is typically called the state action requirement. We are protected, for example, against

---

22. *See* Meyer v. Nebraska, 262 U.S. 390 (1923) (state law prohibiting the teaching of any modern language other than English unconstitutionally infringes freedom of parents to oversee children's upbringing and education).

23. *See* San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 22-24 (1973).

24. *See* Roe v. Wade, 410 U.S. 113 (1973).

25. *See* Harris v. McRae, 448 U.S. 297 (1980) (federal law withholding funds for even medically necessary abortions upheld; no general constitutional right to an abortion, only right to contract for abortion free of state interference); Maher v. Roe, 432 U.S. 464 (1977) (equal protection clause does not compel state to pay for medically necessary abortions although state may pay for indigent women's childbirth expenses).

26. Rust v. Sullivan, 111 S. Ct. 1759 (1991).

the state's censorship of certain ideas or modes of expression. Were a state to criminalize the utterance of communist, atheistic, Catholic, feminist, or white supremacist beliefs, such a statute most certainly would be ruled unconstitutional. We are not protected, however, against censorship of those same ideas by private publishers.[27] Should the major publishers determine that certain ideas—communist, feminist, pacifist—do not sell and, therefore, decide not to publish, or should the media decide that certain points of view—critical perspectives on the Persian Gulf War, for example—decrease ratings and, therefore, decide not to air them, effectively censoring from the public discourse those contributions, there has been an unquestionable censoring of ideas from the public sphere. Nevertheless, there has been no constitutional violation.[28] In fact, according to some commentators, Congress's attempts to correct for this private censorship and impose upon private media obligations of fairness may be a constitutional violation of the private media's right to uncensored expression.[29] Consequently, while we all are protected against a wide range of official state censorship, women are not protected against the censorial, silencing effect of a pornography industry run amok,[30] and African-Americans are not protected against the similarly silencing effect of racist hate speech[31]—the murder of the spirit, to use the expression coined by law professor Patricia Williams.[32] Similarly, while we are constitutionally protected against police violence and brutality, we receive no constitutional protection against violence and brutality from a fellow citizen, an abusive spouse, a lover, or a parent. Of course, the state's criminal law may or may not accord

---

27. *See* Hudnut v. American Booksellers Ass'n, 771 F.2d 323 (7th Cir. 1985), *aff'd*, 475 U.S. 1001 (1986). For a discussion of the injurious consequences of private-market censorship of unpopular ideas, see Louise Armstrong, *Dissent for the Duration . . .: Louise Armstrong Talks to Andrea Dworkin*, WOMEN'S REVIEW OF BOOKS, May 1986, at 5.

28. CBS v. Democratic Nat'l Comm., 412 U.S. 94 (1973) (refusal of CBS to accept DNC's editorial advertisements did not violate latter's constitutional rights). According to the Court in *CBS*, to limit journalistic discretion in the name of First Amendment rights would be "anomalous" and a "contradiction." *Id.* at 120, 121.

29. The Court upheld the fairness doctrine, a complex set of regulations imposing obligations on broadcasters to provide balanced treatment of opposing points of view, in Red Lion Broadcasting Co. v. FCC, 395 U.S. 367 (1969), but the case has generated a vast array of criticism. *See, e.g.*, Kenneth Karst, *Equality as a Central Principle in the First Amendment*, 43 U. CHI. L. REV. 20 (1975); L.A. Powe, Jr., *"Or of the [Broadcast] Press,"* 55 TEX. L. REV. 39 (1976).

30. *See Hudnut*, 771 F.2d at 328-34. *See generally* CATHERINE MACKINNON, *Pornography: On Morality and Politics, in* TOWARD A FEMINIST THEORY OF THE STATE 195-214 (1989) [hereinafter FEMINIST THEORY].

31. *In re* R.A.V., 464 N.W.2d 507 (Minn.), *cert. granted*, 59 U.S.L.W. 3823 (U.S. June 10, 1991) (No. 90-7675).

32. WILLIAMS, *supra* note 19, at 73. Williams is perhaps our only eloquent contemporary poet-lawyer.

us protection from such private violence, but whether it does or not is of no constitutional moment. Even if the state does nothing to protect us against such violence, there has been no constitutional violation. So long as the violence came from a private citizen, there has been no state action. At worst, there has been only state inaction, and that, as the Supreme Court has made clear, is simply not enough. The citizen, we might say, has no constitutional right to a police force.[33]

To recapitulate, the modern concept of ordered liberty governing the great bulk of our modern constitutional law is constituted by, and limited by, two principles: (1) the philosophical and political notion that there is some sphere of individual conduct, belief, and expression that should be inviolable against the intrusion, intervention, or interference of social authority; and (2) the more purely constitutional (and distinctively American) notion that the individual's negative freedom has been infringed wrongly only if it is the state, rather than some other social authority or private force, responsible for the infringement. Before beginning my critique, it may be worth noting one general logical feature of the liberal concept of ordered liberty as I have just described it. Contrary to a widespread misunderstanding, the two principles that constitute and limit the modern understanding of ordered liberty—the preference for negative liberty and the state action requirement—are logically independent of each other. Not only is the state action requirement not required by the preference for negative over positive liberty, but in many cases, it is fundamentally at odds with it. If we are truly concerned with the negative freedom of individuals, then we should be concerned with unnecessary limitations on our interference with those freedoms whatever the source, whether it be the state or some other form of organized social authority. There surely are forms of organized social authority that are at times more intrusive, more interventionist, more controlling, and more interfering with an individual's right to be left alone than the state. Indeed, it may only be through state intervention that these private infringements of the individual's negative liberty can be addressed.

Imagine, for example, the profound interference with the negative liberty to do, think, act, believe and say as one pleases, worked by some Mormon communities on the developing sense of self and society of thirteen- or fourteen-year-old adolescent girls, primed by their parents and their community not for participatory and autonomous adulthood, but for continuing infantilization and dependency through a too-early marriage. Imagine the similar effect on the negative liberty of the Amish child occasioned by the Amish com-

33. *See* DeShaney v. Winnebago County Dep't of Social Servs., 489 U.S. 189, 196-97 (1989).

munity's refusal to allow their children a high school education. There may be reasons, even compelling reasons, for insisting that the state ought not interfere with the practices of these insular religious communities. We may value religious diversity in these subcommunities for their own sake, as Mill urged that we should.[34] Alternatively, we may fear the sort of spill-over consequences of unleashing state power on such groups. I am not arguing for greater control of these religious minorities. What I insist on is the more limited point that whatever the argument might be for nonintervention by the state into the freedom of these religious groups to oppress their individual members, it cannot be based on the principle of negative liberty standing alone, for that principle often will cut the other way. Although the negative freedom of *groups* or subcommunities to be left alone might be furthered, the *individual*'s negative freedom is hurt, not helped, by the state's policy of nonintervention into these private spheres of communal coercion, intimidation, and control.[35] True devotion to the principle of negative liberty should sometimes counsel for state intervention into private relations and sometimes counsel against it. There is no necessary connection between the respect for individual autonomy, which informs our commitment to negative liberty, and the fear of excessive state control, which informs our constitutional state action doctrine. Both commitments might be justified, but they must be justified on independent grounds; neither follows from the other.

II. Ordered Liberty and Women's Rights

Whatever its internal logic, the modern conception of ordered liberty currently guiding constitutional law has not served women well. The reason is simple enough: the modern conception of ordered liberty does not capture and, so long as the modern interpretation dominates, the Constitution does not guarantee the liberties that women peculiarly lack in this country. As a consequence, the constraints under which women distinctively live are not those prohibited by constitutional mandate. In a formal sense, the problem is twofold. First, many of the liberties women lack are positive rather than negative and not protected for that reason. Second, whether characterized as positive or negative, the constraints that limit women's liberty typically are not imposed by the state, but by private and

34. Mill specifically defended the polygamous practices of the Mormons on just these grounds, but he did so without considering, and perhaps not noticing, that those practices endanger the very individual liberties specifically defended in near absolute terms in earlier sections of his famous essay. *See* Mill, On Liberty, *supra* note 16, at 73.

35. We unduly flatter these spheres with the appellation "community."

sometimes very private, even intimate, relationships and are not prohibited for that reason. Indeed, the mismatch between the liberty protected by the Constitution and the liberty women distinctively lack is so great as to make the Constitution irrelevant at best and often a positive danger to women's lives. From the perspective of women's liberty, it is truly not clear at this point in our history whether the Constitution and the ordered liberty it protects is worthy of celebration or a part of an immense societal problem that still remains to be solved.[36]

I will give two examples of the general incompatibility of women's needs and our ruling, liberal conception of ordered liberty. If we look directly at contemporary women's lives, we can identify two constraints within which women live quite distinctively and which disproportionately limit our freedom. First, women, far more than men, live within the constraints of gender roles assigning to women far greater responsibility for child-raising and domestic labor.[37] This is what Arlie Hochschild provocatively calls the "second shift" phenomenon: women, in effect, work two jobs in this society to a man's one. One of these jobs is often underpaid, and the second, the domestic shift, is utterly unpaid.[38] Consequently, by virtue of their unequal responsibility for domestic and child-care labor, women find it difficult or impossible to be economically self-sufficient through participation in the paid labor market or to be involved in the public sphere of political decisionmaking. There are a limited number of hours in a day, and so long as women continue to work two jobs to a man's one, and continue to be trained to willingly accept this inequity and men trained to expect it, women will find it proportionately more difficult than men to live otherwise autonomous, politically engaged, economically self-sufficient lives. As long as there is laundry to wash, diapers to change, children to feed, houses to clean, and meals to make, and as long as women disproportionately are doing it, there is that much less time for women to vote, campaign, hold public office, sit on boards, create art and culture, and live otherwise positively free lives.[39] Just as important, so long

---

36. That is, if women are ever to be men's equals in the civic, economic, and private spheres in which we live out our lives.

37. *See generally* ARLIE HOCHSCHILD, THE SECOND SHIFT: WORKING PARENTS AND THE REVOLUTION AT HOME (1989); SUSAN MOLLER OKIN, JUSTICE, GENDER AND THE FAMILY (1989); Richard Delgado & Helen Leskovac, *Review Essay—The Politics of Workplace Reforms: Recent Works on Parental Leave and a Father-Daughter Dialogue*, 40 RUTGERS L. REV. 1031 (1988); Nancy E. Dowd, *Work and Family: Restructuring the Workplace*, 32 ARIZ. L. REV. 431 (1990); Jana Singer, *Women's Work*, REP. FROM THE INST. FOR PHIL. AND PUB. POL'Y no. 1, 11 (Winter 1991).

38. *See* HOCHSCHILD, *supra* note 37.

39. For an eloquent treatment of the conflict between mothering and the production of culture, see TILLIE OLSEN, SILENCES 203-12 (1978).

as women are and feel responsible for these tasks, the absolutely obvious incompatibility of that work with the positive liberty praised by classical liberals[40] and modern civic republicans[41] alike—full, rounded, independent, politically participatory lives led in the public sphere—will continue to imply for and to women the inescapable message that we are unsuitable for the liberty men expect and often (but not always) receive as a matter of course. Indeed, as suggested by the sixties term "Women's Liberation," the fact that women find political participation and economic self-sufficiency a much more illusive goal than men might be described as the most important finding of the second wave of twentieth-century feminism that captured our collective political imagination in the 1960s and early 1970s.[42]

Second, women live within the constraints of a high risk of sexual violence and a pervasive fear of sexual violence inhibiting our actions in the public world and coloring our inner lives in the private.[43] This greater vulnerability obviously compromises women's physical security and psychological well-being in many ways of which I will mention only a few. First, both the violence itself and the fear of sexual violence quite obviously and dramatically limit women's freedom to move about physically in our community to a much greater extent than such a fear limits men.[44] Second, sexual violence and fear of sexual violence also drastically limit our choices and even our perception of our choices of ways to live.[45] It makes marriage appear to be much safer and, hence, more desirable than it is. It makes nonmarital life styles—single, celibate, lesbian—both appear to be and in fact to be quite dangerous to say nothing of socially unacceptable. Third, sexual violence and the fear of it limit many women's enjoyment of sexuality, and this, too, should be understood as a very real cost. Most damaging, however, fear of sexual violence, like fear generally, infantilizes women and leaves us more vulnerable, both in our own perceptions of ourselves and in others' perceptions of us. The fear, as much as the actual violence, badly cripples women's sense of ourselves and societal perceptions of us as auton-

---

40. *See* JOHN STUART MILL, UTILITARIANISM (1863); MILL, ON LIBERTY, *supra* note 16.

41. Frank I. Michelman, *Law's Republic*, 97 YALE L.J. 1493 (1988); Cass R. Sunstein, *Beyond the Republican Revival*, 97 YALE L.J. 1539 (1988).

42. *See, e.g.*, BETTY FRIEDAN, FEMININE MYSTIQUE (1983); SIMONE DE BEAUVOIR, THE SECOND SEX (1952).

43. *See* CATHERINE MACKINNON, FEMINIST THEORY, *supra* note 30, at 195-214; CATHERINE MACKINNON, *A Rally Against Rape* and *Sex and Violence: A Perspective*, *in* FEMINISM UNMODIFIED: DISCOURSES ON LIFE AND LAW 81-93 (1987) [hereinafter FEMINISM UNMODIFIED]; DIANE E.H. RUSSELL, RAPE IN MARRIAGE (1982).

44. MARGARET T. GORDON & STEPHANIE RIGER, THE FEMALE FEAR (1989).

45. Adrienne Rich, *Compulsory Heterosexuality and Lesbian Existence*, *in* 5 SIGNS: JOURNAL OF WOMEN IN CULTURE AND SOCIETY 631-60 (Summer 1980).

omous, free and independent agents. For women in abusive marriages and intimate relationships, this infantilization and depersonalization is most extreme. In such relationships, sexual violence and the fear of it can strip away virtually all sense of self-possession. The repeatedly abused woman becomes, in fact as well as in self-image, a means, rather than an end, to the fulfillment of another's desires. She quite literally lacks the capacity to be herself when she has been put under the sovereign will of a violent and violence-prone partner.[46] More generally, the fear of the potential for sexual violence from husbands, partners, potential partners, acquaintances, or strangers leaves all women, not just abused wives and rape victims, considerably more vulnerable, more dependent, and more constrained than our brothers, fathers, sons, and husbands.

Both the constraint of unequal parenting and the constraint of sexual violence profoundly limit women's political participation, economic self-sufficiency, physical security, and psychological well-being—or, in a word, women's autonomy. Both constraints limit some central aspect of women's liberty. What I want to show now is that in spite of the tremendous threat these constraints pose to women's liberty, neither of them, given the dominant, liberal understanding of ordered liberty, is particularly vulnerable to constitutional challenge or within the ambit of constitutional concern. Even worse, the societal conditions that facilitate and at times constitute these constraints may have constitutional protection, in the name of protecting negative liberty, against political or legal change. Let me comment on each of these constraints in a little more detail, showing why they are largely unamenable to constitutional challenge and why the social practices from which they arise may even be constitutionally protected.

I start with women's unequal parenting responsibility and the constraint it imposes on women's political and economic autonomy. Whatever else one might want to say about this particular constraint on women's lives, this much is clear: However unjust it may be and however pervasive its restrictive impact on women's potential, given the modern understanding of ordered liberty under the Constitution, the Constitution holds no promise of correcting it for two reasons. The first should be obvious enough from the way I have labelled the problem. The kind of autonomy of which women are deprived by virtue of the unequal distribution and unequal recognition of and compensation for domestic labor is almost paradigmatically positive rather than negative. It is the freedom to live a certain kind of involved, public, political, and economic life, not freedom from any

46. I have explored this in more detail elsewhere. *See* Robin L. West, *The Difference in Women's Hedonic Lives: A Phenomenological Critique of Feminist Legal Theory*, 3 WISC. WOMEN'S L.J. 81 (1987).

particular kind of intrusion. It is the freedom to be, in the fullest sense, a citizen that is threatened in part by women's unequal responsibility for parenting the young. So long as we continue to pledge our allegiance to a Constitution that protects negative but not positive liberty, the tremendous constraints imposed upon women's public lives by their unequal responsibilities for domestic labor will never rise to a constitutional magnitude. Whether or not it is unjust, it is not an injustice for which the Constitution as it is presently understood demands compensation.

Second, regardless of whether the liberty women lack by virtue of unequal and unpaid parenting is negative or positive, women's unequal parenting and domestic responsibility is still largely invulnerable to constitutional challenge because of the state action requirement. One need not be a naive adherent to a falsely innocent conception of the state to infer from the cross-cultural breadth of the problem and transgenerational depth of the problem that the assignment to women of disproportionate child-raising labor, domestic chores, and of a lesser role in public life is made not by any particular state or state official but by a complex, transsocietal, and transgenerational web of shared understandings about the nature of women and men, women's natural capacity for motherhood and disinclination for the life of the citizen, artist, intellectual, artisan, or wage-paid laborer, and men's societal inclination for all of the above and natural disinclination for parenting. We might, for purposes of brevity, call this complex, transsocietal, transgenerational web of shared understandings "patriarchy." My point is that patriarchy, so defined, is not (or certainly is not entirely) a product of state action no matter how broadly we might define either concept. Patriarchy infects not only our laws, but also our private lives and relations. It springs not only from our legal system, but also from our private orderings. Although the state may have from time to time in our history exacerbated it, legitimated it, enforced it, and may in some ways continue to do so, the state did not create patriarchy. For that fundamental reason, simply ending the state's complicity with it will not cure it. Women living in a state whose law is rigorously neutral toward women and men still will find themselves burdened by the inequality and injustice of a private regime of patriarchal control. Women will still find themselves unable to live the positively free life of the citizen because of it.

Simply put, if patriarchy persists at least to some degree and in some of its manifestations without benefit of state action,[47] there simply is no constitutional violation, so long as we understand the Constitution to protect only our right to be free of state intervention.

---

47. In the example of unequal distribution in the private nuclear family of child-raising responsibility, there may well be no state action.

Regardless of whether the unjust distribution of labor and responsibility in the family sphere constrains women's positive liberty to full citizenship and autonomy or women's negative liberty to choose a way of life free of social authoritarian intervention, there is nothing unconstitutional in the injustice. The Constitution is silent on the many constraints, injustices, and inequalities perpetuated on women by the private forces we understand as patriarchal. In short, patriarchy is constitutional to the extent that it is autonomous from state control and creation.

The difficulty goes even deeper, however. Not only is patriarchy not unconstitutional, but, to the degree that patriarchy is woven into the fabric and pattern of our most private intimate lives, it may be even constitutionally protected. The Court has held repeatedly that our negative liberty to be free of state intervention at a minimum contains the liberty to create a private, familial life in whatever way the individual deems best and in line with her own beliefs about the meaning and content of the good life.[48] The central and liberal understanding that whatever else negative liberty protects it must protect the relations of our intimate and familial lives typically is captured in one word: privacy. Because the Constitution protects our familial privacy, it arguably protects our access to birth control,[49] our right to procure an abortion,[50] to attend the private school of our choice,[51] and, in general, to make whatever decisions we deem best about the way our children are raised.[52] That privacy, however, comes with a terrible and often terrifying price to women. If, as a number of feminists now contend, private life is the home of patriarchy[53]—if patriarchal control of women's choices and patriarchal domination of women's inner and public lives occur in the very private realm of home life—then the Constitution, above all else, protects the very system of power and control that constrains us. The complex system of ties peculiarly binding women and not men may be not only not unconstitutional, but positively constitutionally protected. If so, then the Constitution is not only not a shield against

---

48. *See, e.g.*, Griswold v. Connecticut, 381 U.S. 479 (1965); Pierce v. Society of Sisters, 268 U.S. 510 (1925); Meyer v. Nebraska, 262 U.S. 390 (1923). The privacy cases, which protect traditional, patriarchal, familial arrangements (both nuclear and otherwise and with the exception of the abortion decision), all protect private and social practices that have a pronounced negative impact on women's self-esteem, self-definition, and self-worth. *See* MACKINNON, FEMINIST THEORY, *supra* note 30, at 184-95 (similar critique of the Court's privacy doctrine as insulating patriarchy).

49. *See Griswold*, 381 U.S. at 485-86.

50. *See* Roe v. Wade, 410 U.S. 113, 152-53 (1973).

51. *See Pierce*, 268 U.S. at 534-35.

52. *See* Wisconsin v. Yoder, 406 U.S. 205, 230-34 (1972).

53. This is what is meant by the phrase "the personal is political." *See generally* MACKINNON, FEMINIST THEORY, *supra* note 30, at 41, 94-95, 119-20.

injustice for women, but is itself a sword of injustice pointed very markedly at women. It is part of the problem, not part of the solution.

The constraint on women's liberty occasioned by sexual violence, like that imposed by gender roles, is also not amenable to constitutional challenge under current constitutional interpretation. Unlike the constraint of gender roles, sexual violence might be a constraint on negative rather than positive liberty. As was the case with gender roles, however, the constraint on liberty occasioned by sexual violence is not a constraint directly worked by state action. Instead, it is a constraint imposed by men. Although the state unarguably aggravates the harm by casual, lax, or nonexistent enforcement of the criminal laws against sexual violence,[54] it is the sexual violence actually perpetrated by men—strangers, acquaintances, dates, lovers, and husbands—rather than irrational or abusive states or state officials that most profoundly limits women's liberty. To put the same point affirmatively, while we have a panoply of rights protecting us against abusive and violent action by the state, we do not have a constitutional right to be free of sexual violence. Because of the so-called state action requirement, the profound infringement of women's liberty by sexual violence violates no constitutional right of sexual security, invokes no constitutional norm of ordered liberty, and triggers no constitutionally significant obligations. There is simply no real constitutional issue.

Thus, (and this is the central point of my critique) because the fear of sexual violence is not a fear of abusive state action, it is of absolutely no constitutional consequence. In the extreme case, arguably no constitutional guarantee would be breached were the state to cease enforcing entirely its criminal laws dealing with sexual violence. This would be an example of state inaction, not state action; and although it would undoubtedly give rise to constitutional litigation, there would be no clear-cut argument supporting such a challenge. The bottom line is that our constitutional guarantee of ordered liberty—our constitutional right to be free of abusive, irrational, or unnecessary infringement of our individual freedom—is a largely empty promise for women. It addresses what is, at worst, a marginal problem in women's lives and leaves absolutely untouched the most glaring source of bondage.

In the case of the constraints of sexual violence no less than the constraints imposed by gender roles, the problem is not just that the constraint on freedom is not unconstitutional, or put affirmatively,

---

54. A dramatic example is the so-called marital rape exemption, which exempts wives from the protection of rape law and exempts husbands from its reach. *See* Robin L. West, *Equality Theory, Marital Rape, and the Promise of the Fourteenth Amendment*, 42 FLA. L. REV. 45 (1990) [hereinafter *Equality Theory*].

that we do not have a constitutional right to be protected against sexual assault. The problem is deeper than that. If the state were to take affirmative actions to address sexual violence and the violence that women suffer within intimate relations and private homes in particular, such action may itself be unconstitutional or, at least, raise constitutional problems. In the interest of the individual's negative liberty to do, think, speak, and act as he or she pleases, the Constitution generally protects the liberty of the individual against excessive or overzealous criminalization of private life and protects a realm of privacy (typically as co-extensive with family life) within which it is extremely difficult for the state to intrude. I am not saying that it would be unconstitutional for the state or for the federal government to undertake legislative action addressing the problem of domestic abuse or sexual violence. If it should do so through criminalization, however, both the general concept and the particular conception of the Constitution as the guardian of individual liberty against the criminal arm of the state would burden and limit its efforts.

The general problem, of which gender roles and sexual violence are but two examples, is that the modern Constitution, in the name of ordered liberty, defines, insulates, and then protects a realm of individual privacy within which the state may not intrude. It is within that very realm, however, that the subordination of women through violence and the threat of violence, through the assumption of unequal parenting obligations, and through the imposition of restrictive gender roles occurs most egregiously. We are left with this uncomfortable and possibly life-threatening constitutional paradox. The Constitution protects and guarantees ordered liberty, but it does not secure women's liberty. The Constitution protects the individual against abusive and violent state conduct, but not only does it not protect women against the abuse and violence that most threatens them, it perversely protects the sphere of privacy and liberty within which the abuse and violence takes place.

The deep incongruity between our modern liberal conception of ordered liberty and women's needs does, of course, have historical parallels. Throughout history, in fact, feminists have felt ambivalent about the Bill of Rights—from Abigail Adams' futile attempt to urge her husband to include women's interests,[55] if not rights, in the early drafting of the original document to the late nineteenth-century abolitionist feminists' bitter disappointment with the Reconstruction Congress's refusal to include women's equality in the vision of social

55. Letter from Abigail Adams to John Adams (March 31, 1776), *in* THE ADAMS PAPERS: ADAMS FAMILY CORRESPONDENCE, DEC. 1761 - MAY 1776, 369-70 (L.H. Butterfield et al. eds., 1963). *See also* ALICE S. ROSSI, THE FEMINIST PAPERS: FROM ADAMS TO BEAUVOIR 7-15 (1973).

justice embodied in the Fourteenth Amendment.[56] The modern Constitution, however, informed by the distinctively modern liberal understanding of ordered liberty, does not only ignore women—although it does do that. It positively protects the sphere of privacy, negative liberty, and individual freedom within which women are most vulnerable and within which women are uniquely, individually, and definitively oppressed.[57] Thus, through its commitment to a liberal and modern conception of liberty, the contemporary Constitution not only fails to protect women's needs and aspirations, but affirmatively protects the sphere of privacy and conduct within which women's subordination occurs.

III. LIBERTY, EQUALITY, AND AUTONOMY

There are two possible ways of addressing this conflict between women's needs, interests, and aspirations and our presently dominant, liberal interpretation of ordered liberty. The first of these, which I will call the egalitarian response, has become the near-standard response of feminist constitutional lawyers. It is, I believe, deeply flawed. The second has not received as much development, but may ultimately have more promise.

The egalitarian response begins with this correct and telling observation: The tension between women's interests and the modern interpretation of ordered liberty is not unique to women but, instead, exemplifies a much larger and deeper phenomenon, which is the tension, conflict, and contradiction between our constitutional commitment to liberty on the one hand and our political commitment to equality on the other.[58] The conflict is not, in other words, between women's liberty and ordered liberty, as I have been describing it, but between equality and liberty. Individual liberty, no matter how construed, always comes at the cost of equality. Individual liberty, so to speak, "frees up" the sphere of action within which private individuals oppress each other. As the New Deal constitutionalists and liberals saw it, "freeing up" individual liberty in the economic sphere exacerbates and exaggerates differences in wealth between owners and laborers. Achieving some more egalitarian distribution of income requires limiting the negative liberty of individual economic actors. Similarly, in our own time, "freeing up" the negative liberty of individuals to say exactly what they please, no matter how racist,

---

56. ELEANOR FLEXNER, CENTURY OF STRUGGLE: THE WOMAN'S RIGHTS MOVEMENT IN THE UNITED STATES 145-50 (1975).

57. *See* Mary Becker, *The Politics of Women's Wrongs and the Bill of "Rights": A Bicentennial Perspective* ___U. CHI. L. R. ___(1992).

58. *See supra* note 3 and accompanying text; *see also* West, *Ideal of Liberty*, *supra* note 11.

hateful, incendiary, or vicious exacerbates the harms of the worst kind of virulent racism still visited upon African-Americans in our society and, consequently, widens the social inequality between the white majority and the black minority.[59] By allowing the individual a free rein in matters of speech, we subject members of racial minorities to injurious, belittling, sometimes emotionally crippling forms of racial insult—speech that we can all agree has absolutely no redeeming value. Likewise, by leaving the individual free to speak, hear, sell, or purchase whatever he or she wishes, we free up the multimillion dollar pornography industry to endanger women's self-image, lives, and safety through violent imagery that arguably increases the risk of sexual violence in an already violent society. By freeing the individual to act absolutely as he wishes within the privacy of his own home, we endanger the well-being and often the lives of children at the whim and mercy of sometimes less than loving parents. Examples could be multiplied.

The lesson to be learned from these conflicts, according to this view, is that increases in individual liberty generally come at the cost of decreases in equality. Put somewhat differently, according to the egalitarian critique, individual liberty invariably exacerbates, rather than ameliorates, the subordination of some groups by others—of women by men, of blacks by whites, of workers by capitalists—and, accordingly, widens the gap in power, prestige, and wealth between these groups. Liberty and equality, on this view, are in an inevitable tension: we cannot increase one without jeopardizing the other. If we want to do something real about equalizing men and women or blacks and whites, we will have to limit, somewhat, individual freedom; and if we want to increase individual liberty, we will have to jeopardize, to some degree, equality. To whatever extent we are constitutionally "constituted by" commitments to both ordered liberty and the civic, political, or, at least, formal equality of men and women, capitalists and laborers, and blacks and whites, we are committed inescapably to contradictory ideals.

The conflict I have been discussing between our modern understanding of ordered liberty and women's needs, on this view, simply partakes of this same general pattern. As noted above, "freeing up" speech facilitates the harms done to women through the propagation of pornography and, thus, exacerbates inequality. Protecting the privacy and freedom of individuals to do and say as we wish in our private, intimate lives frees men to oppress, abuse, exploit or, in the extreme, to rape, and thereby further weaken women. Protecting freedom of speech and expression frees a society riddled by inequities to perpetuate, in the name of freedom of ideas, notions of gender

59. *See supra* note 3 and accompanying text.

roles that continue to impoverish women.[60] Conversely, each gain in gender equality, like gains in equality generally, comes with the price of a diminution in individual freedom: a shrinking of First Amendment freedoms in the case of pornography, a piercing of family and individual privacy in the case of domestic violence and inequitable allocation of responsibility for parenting, and a diminution of individual liberty in the case of greater criminalization of sexual violence and greater enforcement of the sanctions already on the books.

If this general political and philosophical point is right, then the constitutional strategy we should embrace to address the ill fit between our constitutional commitment to ordered liberty and women's needs seems clear enough. Advocates for women's interests should urge a general constitutional right to equality and then argue that the right to equality is of greater magnitude than the countervailing right, with which it is in tension, to individual liberty. If women are guaranteed equality, if not through the failed ERA then through the equal protection clause of the Fourteenth Amendment, at least our commitment to liberty is limited by this counterbalancing commitment to equality. The Constitution, on this view, gives weight to both values, which are concededly in tension and which, accordingly, must be weighed and balanced against each other by a court or interpretive body sensitive to both. The Constitution, therefore, not only protects the individual's negative liberty to speak and to privacy, but also protects women's right to equality. Hence, limits on pornography may be not only constitutional, but constitutionally required. Similarly, I have argued elsewhere, the so-called marital rape exemption—which provides that nonconsensual sex within marriage is not rape and which is still in force in a number of states—may be unconstitutional in spite of the infringement on marital privacy and individual liberty that the criminalization of marital rape entails.[61] Lastly, if this view is right, when individuals arrange their private affairs so as to allocate to women a grossly disproportionate amount of the unpaid and under acknowledged labor of raising the next generation, we face a problem of constitutional, not just moral and political, magnitude.

I am in complete sympathy with the goal of women's equality and also have considerable sympathy for the particular arguments summarized above. There are, however, serious problems with the general conception of the Constitution on which these arguments rely. First, as a doctrinal matter, it is not at all clear that the

---

60. *See* MACKINNON, FEMINIST THEORY, *supra* note 30, at 195-214; MACKINNON, FEMINISM UNMODIFIED, *supra* note 43, at 184-94.

61. *See* West, *Equality Theory*, *supra* note 57; Note, *To Have and To Hold: The Marital Rape Exemption and the Fourteenth Amendment*, 99 HARV. L. REV. 1255 (1986).

Constitution contains even a general commitment to anyone's equality, for women or for any other group. The Fourteenth Amendment does, of course, guarantee us equal protection of the law, but it is unlikely that the framers intended the clause to mean, or the Court will ever interpret it as meaning, that the Constitution requires the sort of social, political, and economic equality women lack and that is threatened by an unbridled liberal devotion to ordered liberty.[62] As a purely strategic or prudential matter, then, the attempt to balance the commitment to liberty with a countervailing commitment of an equal constitutional magnitude to equality seems doomed to failure. There is no constitutional commitment to equality that comes anywhere near the weight, depth, breadth, history, or sincerity of our constitutional commitment to liberty. In any constitutional standoff between liberty and equality, liberty is going to win. Liberty is an unmistakably constitutional requirement as well as a political and moral aspiration, while equality is, at best, a political aspiration and, at various points in our history including most notably this one, not a widely shared one.

The more basic problem, however, with this liberty versus equality view is that by conceiving of the needs, interests, and aspirations of women that are threatened by ordered liberty—interests in security, needs for economic self-sufficiency, and aspirations for cultural and political participation—as being symptomatic of inequality, egalitarians may have misdiagnosed the problem. The sorts of needs and interests at stake in these conflicts seem to be interests in, needs for, and aspirations of liberty, not equality.[63] Women need to be free of sexual violence both in the home and out in order not only to be equal, but also in order to be *free* in the most basic sense in which that ideal is ever invoked—to have freedom of movement from place to place at the time of one's choosing and for one's own chosen ends. While freedom from sexual violence ultimately would serve to equalize the relative social and economic positions of women and men, it is basically women's liberty and only secondarily women's equality that is lost when women lose the freedom to move about in public spaces free of the fear of molestation. Similarly, women need to be free of disproportionate obligations of labor in childraising not only in order to be equal, but also in order to be free to do other things—to be a fully participatory citizen, to work in the paid labor market, to create art, poetry, sculpture or ceramics, to philosophize,

---

62. *See* Robin L. West, *Toward an Abolitionist Interpretation of the Fourteenth Amendment*, 94 W. VA. L. REV. 111 (1991). It is worth noting that our popular fundamental texts—the Pledge of Allegiance, America the Beautiful, the Star Spangled Banner—make no mention of equality but repeated references to liberty.

63. This was reflected in the since-discarded self-appellation of feminism during the sixties and early seventies as a movement of "Women's Liberation."

educate, or study. Again, freedom from unequal and unpaid child-raising obligations unquestionably would serve to equalize women and men in any number of ways. What each woman loses when she is tied to burdensome and unfair domestic obligations, however, is not simply some share of an abstract group interest in the equality of women and men generally, but rather, and again in the most immediate sense imaginable, her own very individual and very personal liberty.

What I want to suggest is that instead of trying to limit liberty by urging equality as a counterweight, we should undertake, instead, a reconstruction of the modern interpretation of ordered liberty presently dominating both doctrine and understanding so as to include the liberties women distinctively lack. The place that reconstruction should start, I submit, is with the possibility that the modern interpretation of ordered liberty as protecting only negative liberty, and then only negative liberty infringed upon by the state rather than by other non-state authorities, is a flawed understanding of our constitutional tradition. The two limitations defining the modern conception of ordered liberty and rendering the Constitution's promise so empty from the perspective of women's lives and needs are flatly unjustified, given the breadth of political vision that inspired the general phrases of the Fourteenth Amendment, including its guarantee of liberty.

Let me begin with the distinction between negative and positive liberty. Whatever may be the merits of Berlin's assessment of the comparative abstract value of negative and positive liberty, it is far more consistent with the abolitionist history of the Fourteenth Amendment to understand the liberty guaranteed by that amendment's Due Process Clause in a positive rather than negative sense.[64] What the post-Civil War reconstruction amendments were about fundamentally, after all, was securing the positive liberties of citizenship, self-governance, autonomy, and the end of bondage for the freed slaves.[65] The war was not fought to ensure the privacy of the slave or to secure his negative right to read, think, act, and speak as he pleased free of state intervention. It just would not have been enough for the southern states to grant the slaves rights of privacy and liberty to read, think, and speak as they see fit yet leave them slaves—nonvoting, dependent, uncompensated, and unfree. In short, the war was not fought nor the reconstruction amendments passed to ensure the negative liberty of the slave. The war was fought (and surely this was primary) to ensure the slave's positive rights to self-governance, independence, autonomy, and full citizenship. The right

64. JACOBUS TEN BROEK, EQUAL UNDER LAW (originally THE ANTISLAVERY ORIGINS OF THE FOURTEENTH AMENDMENT) (1965).

65. *Id.* at 234-39.

of the citizen to enjoy his liberty and the state's obligation not to deprive him of it other than by due process of law guaranteed by the Due Process Clause of the Fourteenth Amendment must be understood as including these positive rights of autonomy, economic self-sufficiency, and political self-governance.

Second, at least judging by the federal legislation passed in their immediate wake, a goodly part of what those amendments were intended to ensure was the positive liberty of the newly freed slaves not just against pernicious state action, but also against pernicious private action, which included the private relationship of master and slave itself, the private lynchings by the Ku Klux Klan, and the private refusals of service by innkeepers.[66] In the post-Civil War era, legislation and other actions taken by the southern states unquestionably endangered the freed slaves. The greatest threat to the slaves and to their very lives was not state action, however, but private action coupled with state inaction, or in other words, the states' refusal to act against life-threatening and highly organized attempts by private individuals and organizations to deprive the freed slaves of their lives and liberty. It was private, not state action, that posed the most immediate threat to both the negative and positive liberties of the freed slaves.[67] Whatever else might be muddied about the intent of the framers of the Thirteenth, Fourteenth, and Fifteenth Amendments, one thing is vividly clear from the Civil Rights legislation and particularly the Ku Klux Klan Act that followed: What was sought by this profound enlargement of our constitutional charter was a guarantee from private violence and private oppression toward the freed slaves. This included private violence facilitated not only by actions taken by the states, but also by the states' inaction, whether by design or negligence, in the face of threats from private forces and individuals to the security of the former slaves' lives and freedoms.

Thus, the most immediate history of the Fourteenth Amendment, which is the necessary constitutional origin of our modern commitment to ordered liberty, is profoundly at odds with the modern liberal conception of the liberty that amendment was intended to ensure. There is no doubt that the reconstruction amendments were intended in part to protect a sphere of negative liberty. By virtue of their slavery, the slaves indeed lacked what we now call the negative liberties of familial privacy: reproductive freedom, control, and re-

---

66. *See* Ku Klux Act of Feb. 28, 1971, ch. 99, 16 Stat. 433, ch. 22, 17 Stat. 13; Public Accommodations Act of March 1, 1875, ch. 114, 18 Stat. 335.

67. Such liberties would include the negative liberty to move freely about or simply to live, to contract to sell or buy property, or to earn a wage for one's labor and the positive liberty to vote, run for office, assume the rights and responsibilities of full political citizenship.

sponsibility; and freedom of thought and religion. There is also no doubt, though, that the reconstruction amendments were intended to wipe out slavery itself and not just these manifestations of it. They were intended to ensure not only these negative rights of free choice and privacy, but also the full positive liberty to which slavery is the absolute antithesis. There is also little doubt that the framers of the reconstruction amendments intended to render unconstitutional a wide range of state actions that were meant to maintain the actual if not the nominal relation of slave and master. But, again, it is absolutely clear not only from the record of the debate, but also by virtue of the wide-ranging legislation that followed their passage, that the amendments were intended to effect far more than pernicious state action. They also were intended to ensure freedom from private violence and oppression and to accomplish this by obligating the states to take affirmative action (to use a modern phrase) to prohibit, penalize, and criminalize (and thereby protect against) private deprivations of that positive liberty. Lastly, according to the explicit command of section five of the Fourteenth Amendment, the amendments were intended to ensure that if the states failed to act accordingly Congress would act in their stead.

Is there any modern lesson for contemporary life to be learned from this history? I think there is. As I have argued above, what women lack most profoundly in this culture is positive, not negative liberty. Women enjoy wide ranging rights to privacy, speech, thought, and religion. What women lack is the enjoyment of positive rights of autonomy, self-possession, economic self-sufficiency, and self-governance, to say nothing of the full rights and responsibilities of citizenship. Furthermore, women lack these liberties not because of pernicious state action but because of widespread and disabling patterns of private discrimination, societal indoctrination, and personal, intimate sexual violence coupled with pernicious or at least negligent state inaction. These conditions appear to be invincible to constitutional challenge. Indeed, to some degree they appear to be constitutionally protected. The consequence is that many women are and feel themselves to be constitutionally disenfranchised.

What I have argued in this paper is that we should be very cautious in identifying the cause of this disenfranchisement as our constitutional history and women's exclusion from it, rather than modern and contemporary understandings. The two limits most modern interpreters read into our conception of ordered liberty—a preference for negative liberty and an insistence on state action—are a product not of our constitutional history, but of modern habits of the heart and mind. In fact, as a matter of constitutional history, the liberal limits we impose on our conception of ordered liberty may be utterly unjustified. If our constitutional history and, hence, our inherited constitutional meanings are broader, more ambitious, and indeed nobler than we have grown to believe, then the disabling

contradiction between our constitutional aspirations of individual liberty on the one hand and our political (whether or not constitutional) aspirations of political equality for women and men on the other may be more apparent than real. If so, then Congress and the states may have an affirmative obligation under the Due Process Clause of the Fourteenth Amendment not only to protect men and women's right to be left alone, but also to protect women against private infringements of their right to be free of sexual violence and to be free of onerous domestic responsibilities that deprive us of full economic and political autonomy. Finally, were Congress and the states to act on these obligations, then women, in spite of our historical exclusion from the process of constitutionalizing and amending this country's foundational beliefs, might come to have what women presently lack—some real stake in the constitutional system of rights and liberties that continues, however imperfectly, to give dignity to us all.

THE COMING AMERICAN WOMAN

EMILY E. SLOAN*

Years ago I formed the habit of reading Dr. Crane's[1] editorials, and I found many a bit of wisdom compactly set forth in words easy to remember. I did not agree with everything he had to say, and there were times when it struck me that he thought he knew a whole lot, but on the whole I found his articles pleasant, and often helpful. I think I was something of a disciple of his until one day I ran across an editorial entitled "Women," and his concentrated wisdom burst forth in this generalization, "Women are divided into two classes, the Marys and the Marthas," and then followed one of the most absurd articles I ever chanced to read. I was both amused and indignant, and found myself saying with some ancient oracle, "Ye gods! Must we endure all this?" He might just as well have said, "Women are divided into two classes, the sheep and the goats." The sheep, of course, meaning those who followed their shepherd or master, and the goats being those who took a notion once in a while to break away from the main herd, and thereby got the worst of it. This man creating this article did what the average person does who tries to further a project strictly on his own. He told simply a half truth, and the master mind who inserted in our old copybooks of the 1880's, the maxim, "Half the truth is often a great lie" hit the nail on the head, smack!

Behold the man who dwindles us down to Marys and Marthas, has forgotten the Mauds, Margarets, Myrtles and Mirandas, to say

---

* Emily Sloan became a lawyer in 1919. She lived and practiced law in Billings, Montana, from 1919 until 1939, with brief exceptions. In 1924, Ms. Sloan became the first female county attorney in Montana. For a slightly longer biography of Ms. Sloan, see Bari Burke, *Afterword: Pulling for the Shore of Independence*, 59 TENN. L. REV. 479 (1992).

"The Coming American Woman" is an address Ms. Sloan delivered to the Y.P.L. of St. John's Lutheran Church in eastern Montana, sometime during the time she lived there. The occasion and audience of Ms. Sloan's address may explain the religious allusions in her address. Ms. Elsie Amlong, Ms. Sloan's eldest daughter, preserved the original, typewritten version of the address. Ms. Amlong kindly granted permission to publish *The Coming American Woman*. Professor Burke added the footnotes to Ms. Sloan's essay.

1. "The women's magazines of the 1920s deluged mothers with advice on raising children." DOROTHY M. BROWN, SETTING A COURSE: AMERICAN WOMEN IN THE 1920s 118 (1987). Dr. Frank Crane was one of the physicians who wrote such advice columns for mothers. *Id.* at 119.

nothing of the Eves who are still abroad in the land, with a few Cleopatras, Ruths and Jezebels thrown in for good measure. Never think for a moment that any man, no matter how great, can line us up in two columns and give a brief summary upon womankind that will include us all. In the words of the Wyoming sheepherder, "It can't be did."

The commotion at the present time is, we are somewhat new in politics, having had less than half a century experience in that line. So we are still in the process of evolution. One of my fellow lawyers said one day, some twenty-five years ago, that he thought, considering the progress we have made during the last decade, after being down and out ever since Eve played hookey, that we are going to wake up some fine day to find ourselves where we honestly belong. Really, girls, that man has what I call Faith. I would there were more of his caliber.

When I started out building a new world for myself, I thought of other women only in the abstract, and it isn't to be wondered at, I had been marooned from women and everything that the average woman holds dear, for so many years that my whole thought was, "I must extricate myself before I can help any one else."

I truly meant to keep still and not express any opinions either wise, or otherwise, but I was asked on every turn by people of both sexes, and in all walks of life, just what I think is ahead for us women, and more particularly what I think the coming American woman will be. So, after many trying experiences and much meditation, I give you my opinion for what it is worth.

The subject is so high, so deep, so broad, that I almost wonder at myself for presuming to answer the questioning public.

In the beginning, that is, the beginning of American history, the Pilgrim dads and the Revolutionary Pilgrims made a vast and mighty mistake, and we women have been paying for it ever since. Why they thought for a minute they were doing all the pioneering, Heaven only knows.

In those good old days in the average run-of-the-mill family the daughter or daughters were taught to do all kinds of housework, cooking, scrubbing, sewing, knitting, weaving, milking, churning, taking care of babies, just to name a few of the more essential duties of the life of ordinary womanhood. She was bossed by her parents and if she married she was bossed by her husband. A man seemingly didn't think he was much of a man if he didn't have a woman to do his bidding, and if a girl didn't marry she was the object of pity or of scorn in the neighborhood in which she lived. You may have heard of a certain little girl who looked up from her play one day and said to her mother, "Say, ma, if I grow up and get married will I have to marry a man like pa?" and her mother said, "Yes, dear." "And say, ma, if I grow up and don't get married will I have to be an old maid like Aunt Kate?" Again the answer, "Yes,

dear." "Well, ma, it's a mighty tough world on us women folks, ain't it?"

The other type family, those who could afford to spend money on the daughters, brought up each girl in the firm belief that her chief object in life was to allure and charm each and every young man she became acquainted with, until she captured a husband, and then — but why speak of it? Who ever heard of her afterward?

Why boys should have all the fun nobody knows, but the girl was denied all the natural romping fun of girlhood, and was either a doll or a dummy set up for man's admiration, and if he failed to admire she was left alone and lonely. If she had been fortunate enough to inherit money from her parents, all well and good; she might live, after the old folks died, with any condescending relative who would have her under his roof. At twenty-five she was an old maid, at thirty her doomed was sealed. That accounted in a way, for so many mismated couples. A girl during her teens might pass up one or two young men in hopes that her ideal would finally arrive. By the time she was twenty she began to feel anxious. At twenty-two her anxiety had grown to worry. At twenty-five the worry was replaced by desperation. What else could she do, except marry? It was not ladylike for her to earn her own living, so there was nothing left for her, absolutely nothing at all, because if she did not marry, the unenviable life of an old maid stared her in the face. Her lot would be to sing lullabies to her sister's babies, to darn her brother-in-law's sox; to do all the odds and ends in a home not her own; to be grudgingly included in married women's parties, and completely ignored by the young. If she had an idea of her own, people laughed or were shocked, according to their different natures. Of course, there were a few Marys or Joans, such as Susan B. Anthony[2] and Frances E. Willard,[3] but they were exceptions. Remember that.

With the bachelor girl from the masses, her fate was practically the same, excepting of course, she had to depend upon the charity of her nearest relative, and to be forever twitted by the small fry because she wasn't married.

Then girls began to change. Somebody condescended to allow them to teach school. Men had more important work to do. So girls

---

2. Alma Lutz, *Susan Brownell Anthony in* NOTABLE AMERICAN WOMEN 1607-1950, 51-57 (Edward T. James, et al. eds., 1971).

3. Frances Elizabeth Caroline Willard (September 28, 1839-February 17, 1898) was a prominent educator, feminist, and suffragist. She was best known, however, for her long and effective tenure as President of the Women's Christian Temperance Union from 1879 until her death. Mary E. Dillon, *Frances Elizabeth Caroline Willard, in* NOTABLE AMERICAN WOMEN, 1607-1950, 613-618 (Edward T. James et al. eds., 1971).

began going into the back woods and out on the prairie and wrangling the young through the intricacies of arithmetic and all the mysteries of grammar. This went on for a decade or more. Then the mannish girl appeared. She hadn't much of an idea as to what she was going to do, but one thing she knew definitely. She was tired of being dictated to. She wanted her independence. She balked at the word OBEY in the marriage ceremony. She began to invade the places of business hitherto held exclusively by men. They didn't like that a little bit. They were willing that girls should perform the underpaid tasks which men despised. They might even help them in securing such positions. Then we heard a lot of twaddle about where woman's place was. Seemingly just where man wanted to put her. There were the suffragettes and the anti-suffragettes, and the antis were very strong in declaring the world was going to the bow-wows on the run, now that woman was forsaking her home and the fireside, and they were very sure she was forsaking both.

We had no end of newspaper jokes about dad rocking the cradle, and mother donning the trousers and smoking cigarettes. What kind of woman was it, anyway, who refused to marry the first man who asked her, or if she did marry him, refused to look upon him as a master?

In the beginning, and I mean way back in the Garden of Eden, it was decided that it was not good for man to live alone. Adam was not given either a doll or a servant, but he was given a very live partner and a mate. Well, time went on, and just when man started to be lord of all creation, including woman, nobody knows, but the servant problem was presumably responsible for this. Adam tried to help a little but he was of single mind. He was utterly incapable of boiling the kettle and tending the baby at the same time, so he said, "You are much more competent than I am, Eve. See, the baby howls for you. I think my whiskers scare it. Anyway, I saw some of the dandiest speckled fish in the brook yonder, and I think if I brought some of them for little Cain to eat, and a few lion cubs and a parrot or two for him to play with, he might behave himself and not make Abel and the baby squall so much; and it will make things easier for you, dear. And say, Eve, you look awfully sweet with roses in your hair, and I'll just go through the jungle and bring you back a few, and besides, sweetheart, this hut is beastly hot; it's better for my health outside."

And Eve, who had held her own with Adam up to that time, wavered when she thought of Adam's health, and fell for the decorations and the sweet words. We all do. It's so nice to be loved.

But the roses got to be an old story, and the lion cubs didn't improve Cain's disposition, and she grew tired of frying fish, so she decided that a swim in the surf might be a little fun. So she put the baby in a corner, left the boys playing with the lion cubs, and started out through the jungle herself. She hadn't gone far when a wail

pierced the air, a long drawn out pitiful wail. Eve hesitated. The wail grew in volume. All three of them had discovered her departure. She turned and made her way swiftly back to them. That night she explained the matter to friend husband. He told her he thought it was awfully sweet of her to come back. He admired her for her self sacrifice. As for him, he simply could not endure the cry of an infant. It drove him insane. He simply had to take to the bush. He did wonder if those kids would ever reach maturity. It seemed to take them much longer to grow than it did the lion cubs. And he wondered if, when they did grow up, they would keep up that everlasting howling.

When the daughters grew up they naturally followed what they saw their mother do. They didn't know any other way. Yet all through the ages we have found women who simply could *not* follow the beaten trail. Solomon, in all his glory, complained bitterly in regard to women who would speak their minds, and Paul went farther, being a bachelor, and told the women to keep still, and even advised them to keep their heads covered. They might not be so noticeable that way.

Then we hear how man usurped the trousers. It is popularly supposed that trousers were always man's own garment, but it is not so. His first step after wearing a ruffle around his hips was to wear a flowing robe. Woman wore trousers first, and man took them away from her. He by that time had grown to be the master and she the servant. It happened thus: In a certain Arab village where woman was clothed in a garment which was a cross between a skirt and a pair of trousers, and her lord and master, the honorable husband went around in flowing robes. Then one day this husband got into trouble, and he found that his enemies were close on his trail. In fact a servant hurried to him with the news that the enemy was already within the camp.

What should he do? He looked this way and that. There was no hope of escape, and their numbers were too great for him to face alone. Suddenly he had an inspiration.

"Oh, Martha," he cried, "Give me that peculiar pair of pajamas thou wearest, and take this robe and wrap it around thee, and go into the fartherest tent and wait, and when mine enemies come and seize thee, and uncover thy face, behold, they shall turn thee loose, and do thee no harm. And when they see me in these strange trouserines, they will say, 'Alas, nobody but women inhabit this camp,' and they shall shake the dust from their feet and hit the trail."

And Martha being a dutiful wife, gave over the desirable garment and took the robe. The ruse was successful, and the disappointed enemy departed. Then faithful Martha came back to don her rightful garment, but behold his serene highness said to her, "I have tried out these pantalets and found them comfortable. I am not hampered

when I walk. Yea, I can even run with these independent feeling articles upon my limbs; whereas, before, I have been greatly hampered with the wind wrapping my voluminous robes about my shins, and have been scarcely able to navigate at all. Wherefore I pray thee, Martha dear, give me these comfortable duds. Thou dost look so graceful in that robe. In fact, thou dost look queenly, whilst these ungainly things make thee appear unattractive to man."

Poor Martha faltered. "I can scarce carry water and gather figs whilst arrayed in these robes," she objected timidly.

"Martha, darling, I will help thee." And Martha was innocent enough to believe him, and parted with the beloved pants. Therein was her downfall, and the downfall of her daughters and her granddaughters. Man had learned the subtle art of flattery and women fell for it right and left.

They dressed for admiration, and they dressed to outdo each other. There was very little difference in the wealthy women from generation to generation, and the poorer classes mimicked the wealthy in so far as they were able, which by the way, wasn't very far, until after the days of the spinning wheel had passed.

When the plantation days of the South began, the woman of America started trying to live up to whatever her idea of a lady happened to be. She had grown tired of being a drudge. She didn't think anything about the ballot at that time, but she did think of higher education. She wanted to know things. The little girl began to wonder why it was that her brother could be sent to an academy in some distant town to further his education while she was left to gather what she could in the little red schoolhouse and was supposed to be satisfied with what she obtained there. There was a long, hard fight before she entered college.

The next step she took was learning to be charming as well as useful. She carried a load of unnecessary clothing. She pinched her waist and feet, and she learned to swoon at the slightest provocation. And that nonsense kept up for a generation or two. Why, the heroine in the Civil War novel simply had to swoon once or twice in every chapter, or she would not have been up to date. And that was one cause that awakened a few women and put them to work. The condition had to be changed. The woman had to be changed, both from the useless ornament and the helpless drudge, back to the image of her first mother, who was not Adam's servant, but his partner and friend.

I think men subconsciously drifted into being bosses, and women subconsciously drifted into being servants, and when women woke up, and by the help of some sense, started pulling for the shore of Independence, man sat up and took notice. He had a great and overwhelming fear about that time, that he would be obliged to paddle his own canoe alone. To be sure, he had always bragged

about that being his particular right, but he had towed woman's canoe behind his own for so long, and he was so used to paddling two canoes instead of one, that he very nearly had nervous prostration when woman's boat shot out into the current and bid fair to keep up with his own.

Suffrage is nothing more or less than evolution. It came as naturally as the day dawns. Well, perhaps the Sun of Suffrage did come up like thunder in some localities, and men wrung their hands in some instances, and declared that the home was the foundation upon which the Government was built, and that suffrage would destroy the home, and the Government as a natural consequence would go to smash.

I didn't start out to tell you about suffrage. I simply wanted to call your attention to the fact that in the beginning woman was man's partner and helper, and he hers, and that through the mistakes of the human race, woman has occupied one inferior position after another, until it would seem she had reached the depths.

Yet, all along the line there have been splendid men who have appreciated the idea of partnership with their wives. Such a hard old world it has been for women. Even Martin Luther wrote to his wife not to grieve too greatly because of their little daughter's death, for he said, "You know, Maggie, the world is hard with women." And we wonder as we go through the pages of history how woman kept her cheer, her hopefulness, her faith and her courage.

And now, speaking of the typical American woman, whom certain men in high places have declared is becoming masculine, and sending this land of ours to the eternal bow-wows, I would have you one and all sit up and take notice. If these men are referring to the trousers they belong as much to woman as to man. No doubt her ingenuity designed them in the first place. The howlers also tell us that men are becoming effeminate. I deny that. It does not make a man womanish to be polite, it does not detract a single bit from his manliness to give woman a square deal. And the square deal does not hurt her. It simply makes her more self respecting, more self reliant, more generous and kind.

It made me wonder just a little bit when I received a questionnaire from the Woman's Vocational Bureau, when I saw the question, "What would you advise in regard to matrimony for the professional woman, and particularly for the girl who has received her degree in law?" I can see no reason why, having fitted herself for some profession in life, a woman should be deprived of the sweetest and most blessed of life's experiences. A woman with education could scarce help being a more understanding and sympathetic wife, and a wiser and stronger counselor as a mother.

A certain prominent asked me if I thought it paid to educate my daughter when she later married a farmer. I said, "I certainly do.

No profession in life requires quite so much skill and ingenuity as that of a wife and mother. I have never in my life seen a mother who knew too much."

I liked the expression in that man's eyes then, and when he finally spoke, he said, "Mrs. Sloan, I am inclined to think you are right. I never thought of the matter in that light before."

The normal little girl enjoys much the same sports as the average little boy. And do you think she changes fundamentally as she begins to mature? Not so. Every man has a heroine in his heart. Every woman has a hero in her heart. Each tries to live up to what that ideal requires. I mean people who are wholesomely decent. The common plane is found, perhaps more often than we think. Because the great American man has advanced, his sweetheart had done likewise.

So, what is the coming American woman to be? She has been featured in all manner of styles. Some men, who still consider themselves the lords of creation, seem afraid. Some women, to whom progress is too great an effort, shrink from the coming American woman. We need not be affected by the fears of either sex, for there is no cause back of it.

"A little knowledge is a dangerous thing," some sage declared ages ago, and I say, An abundance of knowledge throws open the windows of the soul. When one's mind is broadened there must of necessity be progress, and progress means better things, not only for those who have accepted progress, but for those who come in contact with the progressive.

During my life on the prairie I came in contact with conditions which changed the whole current of my life. In the old days, before the telephone and automobile, when a ranch was an isolated place, the conditions were such that caused the downfall of many a home. Two or three years of that kind of life were enough for the average woman. Men and livestock thrived. Women grew lonesome for their own kind. Solitude did strange things for these women. After the first novelty wore off, a woman could not help growing lonely. I have seen them grow from loneliness to despondency, from despondency to despair. Two routes they usually followed. The first led to a sanatorium; the second, out and away from husband and solitude, with some strange man as guide, for an elopement was not uncommon. I used to be shocked, but now I understand it all.

Progress is curing these evils. For a woman to spend such a life requires a very dull brain or a very quick intellect. The medium type woman was the one who got the worst of the deal all the way through. Her troubles brought me alive, and I am quite aware that we business and professional women are not the first trail blazers, but that we are following the lead of our braver sisters, and are trying to make the path less rough for those who follow after us.

The woman of tomorrow will have all the virtues of the woman of today, plus a more equalized life. She will be physically and

mentally fit to do whatever her hands and her brain find to do, independent, yet comradely and kind. She will be successful as wife, mother, friend, and citizen, because through all the experiences of the ages she is learning to weed out the worst and preserve the best. She will always make mistakes, because she is human and naturally progressive. And progress comes only with effort, and effort follows mistakes, for mistakes are born of ambition, and Woman and Ambition walk hand in hand.

That is why the American woman of tomorrow will not follow the beaten paths of men, but will go with him hand in hand, following some old trails, blazing some new ones. She will be his genuine chum, his companion and helper, just as our much slandered mother Eve was in ages gone by.

We shall not be classified as the Marys and the Marthas. You will see us all, Eve, Rebecca, Ruth and Jezebel; Mary, Lydia, Martha, Cleopatra, Juanita, Marguerite and Susan. Aye, there will be, even as now, a long line of us, and no man will have the temerity to regard us other than as individuals, and if he be wise, he will study the ones with whom he has to deal. He will know better than to flatter Jezebel, and he will keep Eve within sight, lest she find a trail his stumbling footsteps may not follow. He will not upbraid Mary for reading poetry, nor scorn Martha for making sweet pickles. He will hesitate before criticizing Marguerite for using rouge, or Susan for speaking her mind.

He will eventually come alive to the fact that it would be a tiresome old world if we were all Marys and Marthas, just the same as women found out long ago that we didn't want a world composed of Johns and Adonises.

For the most part, man is a contented creature. He likes what he is used to, and Jiggs is a fairly good example. *We* understand the why and the wherefore of Maggie, but lo, these many centuries, we have been keeping these things to ourselves; and though Jiggs may get in a ringer every now and then, and Maggie is perhaps too handy with the rolling pin, still it may be that she is the one from whose arms may come, after much struggling, the future American Woman.

AFTERWORD: PULLING FOR THE SHORE OF INDEPENDENCE

BARI R. BURKE*

> Finally, there is the pervasive theme of "selfhood" for women. The dominant male society suppresses woman's individuality, inhibits her intelligence and talent, and forces her to assume standards of appearance and personality that coincide with the masculine ideal of how a woman should behave and look.
>
> Over the years many feminists have wondered: What would women be like if they were free to develop without being pressured to conform to some pattern set by men? This theme was and remains one that lies close to the heart of feminism; it evokes the visionary image of the "true woman."[1]
>
> Woman's most persistent problem has been to discover for herself an identity not limited by custom or defined by attachment to some man.[2]

Emily Eva Mullenger Sloan's search for selfhood stretched over ninety years and withstood forces that would have defeated less resolute women. During her lifetime, she received few of the customary rewards—recognition, professional respect, social standing, financial security—that the successful exercise of autonomy usually brings for men. Instead, her attempt at autonomy was a daily struggle between her resolve to achieve autonomy, and the social safety and

---

* Professor, University of Montana School of Law. I owe much to my colleagues who are consistently willing to respond to my work, often on short notice. Thank you to Margaret Bentwood, Melissa Harrison, Thomas Huff, Peggy Sanner, Carl Tobias, and Maxine Van de Wetering. I also thank Elsie Amlong, Emily Sloan's oldest daughter, for holding fast to her faith in her mother, knowing that someday someone would come along and celebrate Emily Sloan's contribution to women's legal history. Ms. Amlong is more than a devoted daughter; she is also an incisive editor and a knowledgeable historian.

This essay is part of a sabbatical project researching the lives of women who were admitted to the Montana Bar between 1889 and 1969. I appreciate the generous support of the University of Montana for my sabbatical project. As an outgrowth of my sabbatical work, I plan to write a full-length biography of Emily Mullenger Sloan.

1. FEMINISM, THE ESSENTIAL HISTORICAL WRITINGS xvi (Miriam Schneir ed., 1972).

2. CAROLYN G. HEILBRUN, REINVENTING WOMANHOOD 72 (1979).

positive regard that went with conventional "womanhood."[3] Because Emily Sloan remained a woman despite her characteristically male ambitions and accomplishments, she could never quite measure up in the man's world in which she lived and worked. She came to recognize early that male and female worlds were sharply distinguishable, that women and autonomy were rarely paired, and that "the world is hard with women."[4] Near the end of her life both her hope and her heartache fed this remark: "That I have been a howling failure on the surface there can be little doubt, but somewhere deep in my heart is the consciousness that in spite of appearances, and in spite of all the stumbling blocks and blundering, 'All is well.'"[5] Emily Sloan's essay, *The Coming American Woman*, may be her accounting for the conclusion that "all is well," despite her lifetime of "unapplauded achievement."[6]

Emily Eva Mullenger was born on October 27, 1878, the ninth child and sixth daughter in a family that eventually numbered thirteen children. Wisconsin was her birthplace; her father traveled throughout that state selling insurance. Emily loved the creeks, woods, flowers, and sand pits of northern Wisconsin; it was the beauty of her physical surroundings that she most prized and longed for during the rest of her childhood and married life. These physical surroundings gave Emily great comfort while she received little from her parents who were too busy to dote on any one of the thirteen children.

To Emily's dismay, when she was seven, her family moved to Springfield, South Dakota; eight years later, they moved to Belle Fourche, in the northwestern part of the state. Robert Mullenger, Emily's father, was admitted to the Bar of South Dakota in 1890; he was elected county judge (judge of juvenile and probate court) in 1894, the first year that the family lived in Belle Fourche.[7]

Just twelve days shy of her seventeenth birthday, Emily Mullenger married Al Sloan, hoping to escape her inattentive and unaffectionate family. She lived for seventeen years with him on an isolated South

---

3. Carolyn Heilbrun loosely characterizes the distinction between women and autonomy:

> Men have monopolized human experience, leaving women unable to imagine themselves as both ambitious and female. If I imagine myself (woman has always asked) whole, active, a self, will I not cease, in some profound way, to be a woman? The answer must be: imagine, and the old idea of womanhood be damned.

Id. at 34.

4. Emily Sloan, *The Coming American Woman*, 59 TENN. L. REV. 469 (1992).

5. Emily Sloan, Unpublished Manuscript 3 (approximate date mid-1960s)(copy on file with author).

6. CAROLYN G. HEILBRUN, WRITING A WOMAN'S LIFE 26 (1988).

7. Ms. Sloan's father died during his third term as county judge, while she was pregnant with her fourth child and before she began studying law.

Dakota ranch, raising four children: Edith, Elsie, Dean, and Stanley. Except for the pleasure of her four children, those prairie years were hard on Ms. Sloan: her husband typically ignored her wishes, treating her less as a partner than as his assistant. For example, early in their marriage he filed a homestead claim without first showing the land to Ms. Sloan. Not only was the land desolate, the homestead cabin was a tarpapered shack, ten feet by twelve feet in dimension. Describing her years on the ranch, Ms. Sloan said, "Oh, yes, the prairie's wonderful, for cattle, coyotes, rattlesnakes, and men; but not for women and children."[8]

When Ms. Sloan's children were practically grown, she considered resuming her education and attending college. Although she had wanted to be a writer since the age of eleven and was the author of a volume of published poetry by the age of thirty-one,[9] her husband refused to pay for "a literary course" and convinced Ms. Sloan to begin studying law by enrolling in the American School of Correspondence. She compromised her preference in order to equip herself to escape the ranch someday.

> I told him that all I wanted was to write. I would study anything in that line I could get my hands on. He said if I thought he would be damned fool enough to give me money, or spend his money on anything like that, I just didn't know him. Well, I knew him, all right. He talked a whole afternoon. Finally, when I could endure no more I said, "O-kay. I'll take the law. If I can't have a whole loaf, I'll take a half of one, and convert it into what I want."[10]

Although Ms. Sloan found the law books boring, she completed the correspondence course, earning good marks along the way.

One seemingly trivial event finally convinced Ms. Sloan that her marriage was past hope. Accompanied by her eleven-year-old son, she drove a spring wagon[11] five hundred miles to fetch Elsie who was visiting Edith and Edith's first child. Ms. Sloan asked her husband to look after things at the ranch, including some flowers she had planted to welcome Elsie home from school. He neglected even this small favor.

> It was then something inside of me gave up completely in regard to ever establishing a real home with him. He had no conception of what the word meant. I had so wanted Elsie to see the dooryard pretty and attractive for once in her life. The long years had been a series of letdowns. . . . Well, I had stayed there alone long enough. There comes a limit to human endurance.[12]

---

8. Sloan, Manuscript, *supra* note 5 at 235.
9. EMILY SLOAN, BALLADS OF THE PLAINS (1909).
10. Sloan, Manuscript, *supra* note 5 at 271.
11. Ms. Amlong described the spring wagon as a "double-seated surrey."
12. Sloan, Manuscript, *supra* note 5 at 283, 292.

Ms. Sloan left the ranch when she was thirty-eight "and started out building a new world for myself."[13] On October 1, 1917, she enrolled as a full-time law student at the University of Montana School of Law.[14]

Emily Sloan was admitted to the Bar of Montana on June 10, 1919. She moved to Billings, Montana, where she opened a law practice and made a home. She soon began a political career; in 1920, she ran for the Montana Legislature, but lost in the primary. In 1921, Ms. Sloan lost the race for County Attorney of Carbon County; in 1923, she won that very post by thirty-three votes, thus becoming the first female county attorney in Montana. Two years later, she ran for reelection but lost. A state district court judge in Billings who respected Ms. Sloan's legal work arranged for her to serve as district probation officer, which she did for the following twenty-six months until early 1929. Economic times were notoriously difficult,[15] prompting Ms. Sloan to leave Billings sometime during the early 1930s for Washington, D.C., to look for work with the federal government. She eventually secured a job with the Agricultural Adjustment Administration where she worked for three months before poor working conditions and loneliness provoked her to return to Billings.

She remained in Billings only for a short time before moving to Tacoma, Washington, where her daughters lived. She spent the rest of her life in Tacoma, except for six months (September 1941 through March 1942), when she went to Anchorage, Alaska, to look for work. She held a variety of jobs in Tacoma: she sold cosmetics door-to-door; she worked on a gas rationing board at the shipyards; she worked in the stock room in a drug store; she worked as a clerical worker in an insurance company; she worked in a typewriter supply office; she worked in the Miscellaneous Department at Fort Lawton, soon transferring to the office of the trial judge advocate; she bought and managed an Introduction Club (correspondence club and matrimonial bureau); and she worked two holiday seasons in a large office supply and book store in Seattle.

---

13. Emily Sloan did not divorce her husband until several years later.

14. For Emily Sloan's account of her time studying law by correspondence and at the University of Montana School of Law, see Emily Sloan, *Completing My Education*, 52 MONT. L. REV. 419 (1991).

15. Perhaps the Depression was not the only reason that Ms. Sloan found it difficult to make economic ends meet as a female attorney with her own practice in the 1920s and 1930s in Montana. Catharine Waugh, a graduate of the Union College of Law in 1886, wrote a memoir detailing "[h]er tribulations and frustrations in attempting to situate herself, a single 24-year-old female lawyer, in the male legal establishment of Chicago in 1886." Nancy F. Cott introduces that memoir in an essay that describes some of the obstacles to women's entry into the legal profession. Nancy F. Cott, *Women as Law Clerks: Catharine G. Waugh in* THE FEMALE AUTOGRAPH 160, 161 (Donna C. Stanton ed., 1984).

Not until Ms. Sloan lived in Tacoma for about five years and reached her middle sixties had she saved $125, the cost of applying for admission to the Washington State Bar. The Board of Examiners denied her application on the grounds that she had been out of practice too long. The Board of Examiners granted Ms. Sloan permission to take the bar examination, if she would attend the University of Washington Law School for a few months. Unfortunately, she did not have the money to enroll in school, pay the fee for the bar examination, and open a law office.

Although she never had the opportunity to practice law in Washington, Ms. Sloan continued to write fiction, poetry, and essays, succeeding in having one book[16] and many poems published. In fact, Ms. Sloan won awards for several of her poems.

How could Emily Sloan come to believe that her lifetime of magnificient efforts and accomplishments—raising four children in a one room shack marooned on an isolated ranch, becoming a lawyer in 1919, serving as the first female county attorney in Montana, publishing poetry and fiction—could be perceived as a howling failure? Women learn early that "[i]t's a mighty tough world on us women folks . . . ,"[17] and that sometimes neither of the two available choices (being Mary or Martha, but not Maud, Margaret, Myrtle, or Miranda) suits any one of us. Emily was not a Mary or Martha as her life story discloses.

Although Emily Sloan and Robin West have larger purposes for their essays, both *The Coming American Woman* and *Reconstructing Liberty* are, in part, their authors' versions of the constraints on the female search for selfhood. Professor West and Ms. Sloan agree that the female quest for autonomy conflicts with constraints that "profoundly limit women's political participation, economic self-sufficiency, physical security and psychological well-being . . ."[18] According to West, one of the most effective constraints on women's autonomy is the allocation to women of domestic responsibilities:

> First, women, far more than men, live within the constraints of gender roles [which assign] to women far greater responsibility for child-raising and domestic labor. . . . As long as there is laundry to wash, diapers to change, children to feed, houses to clean, and meals to make, and as long as women disproportionately are doing it, there is that much less time for women to vote, campaign, hold public office, sit on boards, create art and culture, and live otherwise positively free lives.[19]

Ms. Sloan's life story reveals her personal experience with domesticity and childraising: from the age of sixteen, when Emily Sloan

16. EMILY SLOAN, PRAIRIE SCHOOLMA'AM (1956).
17. Sloan, *The Coming American Woman*, *supra* note 4 at 471.
18. Robin West, *Reconstructing Liberty*, 59 Tenn. L. Rev. 441 (1992).
19. *Id.* at 454.

married, until the age of thirty-eight when she left the ranch and began to attend law school, "the constraints of gender roles" trapped Emily Sloan on a barren ranch in South Dakota answering to her husband, caring for her four children, and pining for other people, in particular women ("for three or four months at a time I had not seen a woman's face"[20]). Ms. Sloan published a book of poetry, creating culture, but she had little opportunity to live a public life.

In *Reconstructing Liberty* and *The Coming American Woman*, West and Sloan also agree that patriarchy portrays those gender roles as natural, inevitable, even inescapable. Professor West says:

> [T]he assignment to women of disproportionate child-raising labor, domestic chores, and of a lesser role in public life is made . . . by a . . . web of shared understandings about the nature of women and men, women's *natural capacity for motherhood* and disinclination for the life of the citizen, artist, intellectual, artisan, or wage-paid laborer, and men's societal inclination for all of the above, and *natural disinclination for parenting*. We might . . . call this . . . web of shared understanding "patriarchy."[21]

Ms. Sloan's essay tells the story of how Eve *naturally* became the family caretaker and Adam's subordinate:

> Adam tried to help a little but he was of single mind. He was utterly incapable of boiling the kettle and tending the baby at the same time, so he said, "You are much more competent than I am, Eve. See, the baby howls for you. I think my whiskers scare it. Anyway, I saw some of the dandiest speckled fish in the brook yonder and I think if I brought some of them for little Cain to eat, and a few lion cubs and a parrot or two for him to play with, he might behave himself and not make Abel and the baby squall so much; and it will make things easier for you, dear. . . . [A]nd besides sweetheart, this hut is beastly hot; it's better for my health outside." . . . As for him [Adam], he simply could not endure the cry of an infant. It drove him insane. He simply had to take to the bush.[22]

In addition to the constraint of gender roles, West identifies sexual violence as a fundamental constraint to women's autonomy. Although Ms. Sloan does not mention sexual violence, she repeatedly identifies the techniques of flattery and trickery as part of the process of subordinating women. For example, Eve "who had held her own with Adam up to that time [when he proposed to go catch fish rather than help with child care], wavered when she thought of Adam's health, and fell for the decorations and the sweet words. We all do.

---

20. Sloan, Manuscript, supra note 5 at 283.
21. West, *supra* note 18 at 457 (emphasis added).
22. Sloan, *The Coming American Woman*, supra note 4 at 472-73.

It's so nice to be loved."[23] Next, according to Ms. Sloan, man usurped woman's trousers by telling her that she looked graceful and queenly in his robes rather than her trousers; he also promised her that he would help with the domestic tasks made more difficult by garments that "wrapped around shins" and left the wearer unable to navigate. "Therein was her downfall, and the downfall of her daughters and her granddaughters. Man had learned the subtle art of flattery and women fell for it right and left."[24]

The primary purpose of Emily Sloan's essay was not the same as that of Professor West's essay. West hopes to influence constitutional jurisprudence in a way that protects those spheres of decision-making most central to women's lives;[25] Ms. Sloan hoped to answer the questioning public's concern with the future American woman. Looking not simply to the past to identify the processes of women's subordination, Ms. Sloan shared her vision of the future. Both in her stories of the past and her predictions of the future, Ms. Sloan touched on themes that continue to engage and perplex contemporary feminism: the nature of "ordinary womanhood" and gender role stereotypes,[26] the assignment of domestic labor primarily to women,[27] the institution of motherhood and "the reproduction of mothering,"[28] women as wage-laborers and equal pay for equal work,[29] the possibilities of combining professional work with marriage and motherhood,[30] and the character of relationships between

---

23. *Id.* at 472.

24. *Id.* at 474.

25. West, *supra* note 18.

26. *See e.g.*, SUSAN BROWNMILLER, FEMININITY (1984); CAROL GILLIGAN, IN A DIFFERENT VOICE (1982); CAROLYN G. HEILBRUN, WRITING A WOMAN'S LIFE (1988); CAROLYN G. HEILBRUN, REINVENTING WOMANHOOD (1979); Adrienne C. Rich, *Compulsory Heterosexuality and Lesbian Existence in* POWERS OF DESIRE: THE POLITICS OF SEXUALITY (Ann B. Snitow et al. eds., 1983); Robin West, *Jurisprudence and Gender*, 55 U. CHI. L. REV. 1 (1988).

27. *See, e.g.*, SHULAMITH FIRESTONE, THE DIALECTIC OF SEX (1970); SUSAN MOLLER OKIN, JUSTICE, GENDER & THE FAMILY (1989).

28. *See, e.g.*, DOROTHY DINNERSTEIN, THE MERMAID AND THE MINOTAUR: SEXUAL ARRANGEMENTS AND HUMAN MALAISE (1976); ADRIENNE RICH, OF WOMEN BORN (1976); NANCY CHODOROW, THE REPRODUCTION OF MOTHERING: PSYCHOANALYSIS AND THE SOCIOLOGY OF GENDER (1978); Martha L. Fineman, *Images of Mothers in Poverty Discourse*, 1991 DUKE L.J. 274.

29. *See, e.g.*, Frances E. Olsen, *The Family and the Market: A Study of Ideology and Legal Reform*, 96 HARV. L. REV. 1497 (1983); Roslyn L. Feldberg, *Comparable Worth: Toward Theory and Practice in the United States*, 10 SIGNS 311 (1984); Deborah Rhode, *Occupational Inequality*, 1988 DUKE L.J. 1207; Deborah Rhode, *Perspectives on Professional Women*, 40 STAN. L. REV. 1163 (1988); DEBORAH RHODE, JUSTICE AND GENDER (1989).

30. *See, generally*, FAYE J. CROSBY, JUGGLING (1991); Symposium on Women in the Lawyering Workplace: Feminist Considerations and Practical Solutions, XXXV N.Y.L. Sch. L. Rev. (1990). Women are recapturing our history as well as

women and men.[31] No doubt her wisdom grew from her valiant efforts to extricate herself from the almost total isolation of the male west and to build a new world for herself.

Throughout her life, Emily Sloan rowed toward the shore of independence. She never reached the safety of the shore; she never achieved the luxury of public recognition or financial security. But Emily Sloan apparently knew what Carolyn Heilbrun knows: a woman's exercise of autonomy, the search for female selfhood, is not truly about "happily ever after" endings or finality.

> We women have lived too much with closure: . . . there always seems to loom the possibility of something being over, settled, sweeping clear the way for contentment. This is the delusion of a passive life. When the hope for closure is abandoned, when there is an end to fantasy, adventure for women will begin. Endings . . . are for romance or for daydreams, but not for life.[32]
>
> Safety and closure, which have always been held out to women as the ideals of female destiny, are not places of adventure, or experience, or life. . . They forbid life to be experienced directly. Lord Peter Wimsey once said that nine-tenths of the law of chivalry was a desire to have all the fun. The same might well be said of patriarchy.[33]

Emily Sloan—daughter, sister, wife, mother, lawyer, woman, writer, citizen—exemplifies the "Coming American Woman." I eagerly anticipate the day that more of us follow the lead of our braver sister Emily and try "to make the path less rough for those who follow after us."[34]

---

inventing our future. For a historical account of female attorneys' attempts to combine professional and personal lives, see Virginia Drachman, *"My 'Partner' in Law and Life": Marriage in the Lives of Women Lawyers in Late 19th and Early 20th Century America*, 14 LAW & SOC. INQUIRY 221 (1989).

31. *See generally*, Barbara Easton, *Feminism and the Contemporary Family in* A HERITAGE OF HER OWN: TOWARD A NEW SOCIAL HISTORY OF AMERICAN WOMEN (Nancy F. Cott & Elizabeth H. Pleck eds., 1979); JEAN B. MILLER, TOWARD A NEW PSYCHOLOGY OF WOMEN (1976); Paula Rothenberg, *The Political Nature of Relations Between the Sexes in* BEYOND DOMINATION: NEW PERSPECTIVES ON WOMEN AND PHILOSOPHY (Carol C. Gould ed., 1983).

32. HEILBRUN, WRITING A WOMAN'S LIFE, *supra* note 6 at 130.

33. *Id.* at 20.

34. Sloan, *The Coming American Woman*, *supra* note 4 at 476.

FURTHERING THE INQUIRY: RACE, CLASS, AND CULTURE IN THE FORCED MEDICAL TREATMENT OF PREGNANT WOMEN

LISA C. IKEMOTO*

INTRODUCTION

When the state restricts reproductive choice, it takes control of women's bodies and women's lives.[1] It also expresses at law a reductionist description of female humanhood. By focusing the power of the state on women with regard to their biological capacity to bear children, the law devalues women as persons and describes women as "vessels,"[2] "mother machines,"[3] or "incubators."[4] I too reject as subordinating restrictions on reproductive choice. But in this article I hope to illustrate how describing and addressing the issues surrounding reproductive choice as a gendered issue without regard to race and class precludes us from fully understanding the nature of patriarchy and, in fact, perpetuates it.

More specifically, this article explores the idea that when we view patriarchy from the perspective of gender alone, we take an essentialist position[5] that ranks social and political issues by race and

---

* Assistant Professor, Indiana University School of Law, Indianapolis. L.L.M., Columbia University, 1989; J.D., U.C. Davis (King Hall), 1987. I am grateful to my research assistants, Rolanda Haycox and Jan Dowling.

1. *See* Roe v. Wade, 410 U.S. 113, 153 (1973) ("The detriment that the State would impose upon the pregnant woman by denying this choice altogether is apparent. Specific and direct harm medically diagnosable even in early pregnancy may be involved. Maternity, or additional offspring, may force upon the woman a distressful life and future. Psychological harm may be imminent. . . ."). *See also* Martha A. Field, *Controlling the Woman to Protect the Fetus*, 17 LAW MED. & HEALTH CARE 114 (1989).

2. *Women as Wombs*, Ms., May/June 1991, at 28.

3. *See, e.g.*, GENA COREA, THE MOTHER MACHINE: REPRODUCTIVE TECHNOLOGIES FROM ARTIFICIAL INSEMINATION TO ARTIFICIAL WOMBS (1985).

4. *See, e.g.* Note, *Incubating for the State: The Precarious Autonomy of Persistently Vegetative and Brain-Dead Pregnant Women*, 22 GA. L. REV. 1103 (1988).

5. For broader critiques of essentialism, see Angela Harris, *Race and Essentialism in Feminist Legal Theory*, 42 STAN. L. REV. 581 (1990); Mari J. Matsuda, *Pragmatism Modified and the False Consciousness Problem*, 63 S. CAL. L. REV. 1763 (1990) [hereinafter Matsuda, *Pragmatism Modified*]; Deborah L. Rhode, *The "No-Problem" Problem: Feminist Challenges and Cultural Change*, 100 YALE L.J. 1731, 1786-91 (1991); Susan H. Williams, *Feminism's Search for the Feminine: Essentialism, Utopianism, and Community*, 75 CORNELL L. REV. 700 (1990); Trina Grillo & Stephanie Wildman, *Obscuring the Importance of Race: The Implication of Making Comparisons Between Racism and Sexism (Or Other -Isms)*, 1991 DUKE L.J. 397.

class. As a context for an inquiry beyond gender, I look to cases in which doctors and hospitals have petitioned courts to order the forced medical treatment of pregnant women. It may seem obvious that one must evaluate the forced medical treatment of pregnant women in terms of biologically based gender; courts issue these orders against persons because of their biological capacity to bear children, persons who are necessarily women. It may be less obvious but not novel to explore socially constructed gender,[6] and this article will do so. But if one accepts gender as the only basis for evaluation, one accepts that the dominant culture may prescribe the scope of inquiry.

Two points must be made here. First, the standard legal story opens the door for a discussion of biologically based gender and justifies such distinctions. By responding only to the standard story, we let it dominate the discourse. Elaborating upon gender as a socially constructed, historically rooted category is one way of "seizing the discourse."[7] Second, the standard legal story does not expressly speak to race and class. By failing to look to the experience of women who have been raced and impoverished, we let the standard story blind and silence us. The de facto standard then used to identify, prioritize, and address subordination is the experience of white, middle class women. This excludes and diminishes women of color, particularly those who live in poverty.

Part I of this article sets forth the standard formulae for justifying forced medical treatment of pregnant women and questions some of its basic premises. Part II describes the scholarly response to the standard story and suggests that the standard story, in part, has dictated that response. Part III continues the inquiry into the social construction of gender. In particular, Part III looks beyond the legal and moral formulae to subordination perpetuated on the "local" or social level; it then postulates one feminist response and argues that the feminist response, too, perpetuates patriarchy. Part IV continues the inquiry into race, class, and culture. It explores the silence as well as the words that reveal a mother model premised on negative and positive stereotypes and two claims to cultural superiority. It also suggests a possibility for building choice from coalition.

I. THE STANDARD STORIES

Within the past ten years there have been a number of cases in which doctors and hospital administrators have petitioned for and

---

6. *See, e.g.*, BELL HOOKS, *Feminism: A Transformational Politic, in* THEORETICAL PERSPECTIVES ON SEXUAL DIFFERENCE 185-86 (1990); CATHARINE A. MACKINNON, TOWARD A FEMINIST THEORY OF THE STATE 244 (1989) [hereinafter MACKINNON, FEMINIST THEORY].

7. Linda Greene, Conference on Race Consciousness and Legal Scholarship, Affirmative Action Panel, University of Illinois College of Law, Feb. 22, 1992.

received court orders for the medical treatment of pregnant women without their consent.[8] A 1987 national survey of obstetricians published in the *New England Journal of Medicine* revealed that twenty-one court orders had been sought.[9] The courts issued orders in eighteen of the cases.[10] Most cases simply go unreported. But according to cases described in other medical articles and law reporters, the trend of performing cesarean surgeries, blood transfusions, and other therapies against the woman's will continues.[11]

The legal formulae tell the standard story of court-ordered medical intervention. They speak of rights and interest balancing. The moral formulae tell the same story but speak of social utility and preventing harms. The storytellers, the judge, and those who concur with the idea of forced intervention purport to tell a gender-neutral tale. Critics of the standard story look to the effects of the standard story, indicated by the national survey, and question the claim that these decisions reflect a gender-neutral perspective. They retell the story in ways calculated to reveal that it is in fact gender-biased and explore the implications for all women.

A. The Legal Formulae

The standard legal story, culled from the cases, begins by recognizing the individual's interests in self-determination and bodily integrity.[12] Every competent person, according to traditional liberal analysis, has a right to determine the course of her own treatment, even where the risk of refusing treatment is death.[13] This right may

---

8. The first reported petition for forced medical treatment was denied in Raleigh Fitkin-Paul Morgan Memorial Hospital v. Anderson, 201 A.2d 537 (N.J.), *cert. denied*, 377 U.S. 983 (1964); there are no other reported cases until the Georgia Supreme Court affirmed an order for forced medical treatment in Jefferson v. Griffin Spalding County Hosp. Auth., 274 S.E.2d 457 (Ga. 1981).

9. Veronika E.B. Kolder et al., *Court-Ordered Obstetrical Interventions*, 316 NEW ENG. J. MED. 1192, 1192 (1987).

10. *Id.*

11. The cases cited and discussed in this article may overlap but are not necessarily the same as those identified by the Kolder survey. *See also* Ronna Jurow & Richard H. Paul, *Cesarean Delivery for Fetal Distress Without Maternal Consent*, 63 OBSTET. & GYN. 596, 596-98 (1984) (reports that doctors perform cesareans against the woman's will and without a court order).

12. *See In re* A.C., 573 A.2d 1235, 1243 (D.C. 1990) (en banc); Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130, 1130-31 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986); Taft v. Taft, 446 N.E.2d 395, 396-97 (Mass. 1983); Fosmire v. Nicoleau, 551 N.E.2d 77, 78 (N.Y. 1990). *See also* Union Pacific Ry. Co. v. Botsford, 141 U.S. 250, 251 (1891); Superintendent of Belchertown State Sch. v. Saikewicz, 370 N.E.2d 417, 427 (Mass. 1977); *In re* Conroy, 486 A.2d 1209, 1221-23 (N.J. 1985); *In re* Colyer, 660 P.2d 738, 743 (Wash. 1983).

13. Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2851 (1990); *In re* Quinlan, 355 A.2d 647 (N.J.), *cert. denied sub nom.*, Garger v. New Jersey, 429 U.S. 922 (1976). *But see Cruzan*, 110 S. Ct. at 2859 (Scalia, J., concurring).

be found in common law or in the constitutional right of privacy.[14] In some cases the woman refuses to consent because a blood transfusion[15] or other medical treatment would violate her religious principles.[16] In other cases the refusal expresses fear or other strong emotion.[17] Finally, in some cases the woman may simply disagree with the doctor's assessment of risk or may be willing to take that risk for other nonmedical reasons.[18]

The right to refuse treatment is not absolute. Where there are superior state interests, the state may impose restrictions on medical choice.[19] Those who petition for orders against pregnant women assert as superior the state's interest in protecting the fetus. They may petition to have the court appoint a guardian over the woman,[20] or to have custody of the unborn child placed in the doctor or a state agency,[21] or they may request the court directly to order the woman to comply.[22] But each petition assumes the superiority of the state's interest. The court may express the state's interest in protecting the fetus in a variety of ways. The United States Supreme Court first recognized the state's interest in protecting potential life in *Roe v. Wade*.[23] Some courts simply invoke that interest, as described in

---

14. *Cruzan*, 110 S. Ct. at 2846-47, 2851-52.

15. Jefferson v. Griffin Spalding County Hosp. Auth., 274 S.E.2d 457, 459 (Ga. 1981) ("This refusal is based entirely on the religious beliefs [affiliation unspecified] of Mr. and Mrs. Jefferson"); *Mercy Hosp.*, 489 A.2d at 1131 (Jehovah's Witness); Raleigh Fitkin-Paul Morgan Memorial Hosp. v. Anderson, 201 A.2d 537 (N.J.), *cert. denied*, 377 U.S. 983 (1964) (Jehovah's Witness); *Fosmire*, 551 N.E.2d at 78 (Jehovah's Witness); *In re* Jamaica Hosp., 491 N.Y.S.2d 898 (N.Y. Sup. Ct. 1985) (Jehovah's Witness).

16. *Jefferson*, 274 S.E.2d at 459; *Taft*, 446 N.E.2d at 396 (Born again Christian).

17. *See infra* notes 75-80 and accompanying text.

18. *See infra* notes 137-42 and accompanying text.

19. Cruzan v. Director, Mo. Dep't of Health, 110 S. Ct. 2841, 2852 (1990). The courts have recognized four state interests that may be weighed against a competent adult's right to refuse medical treatment: preserving life, preventing suicide, maintaining the integrity of the medical profession, and protecting innocent parties. *See* Superintendent of Belchertown v. Saikewicz, 370 N.E.2d 417, 425 (Mass. 1977).

20. *See e.g.*, Jefferson v. Griffin Spalding County Hosp. Auth., 274 S.E.2d 457, 458 (Ga. 1981); Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986).

21. *See, e.g.*, *In re* Steven S., 178 Cal. Rptr. 525 (Cal. Ct. App. 1981); Watson A. Bowes, Jr., M.D. & Brad Selgestad, *Fetal Versus Maternal Rights: Medical and Legal Perspectives*, 58 OBSTET. & GYN.. 209, 212 (1981).

22. *See, e.g.*, *In re* A.C., 573 A.2d 1235, 1237 (D.C. 1990) (en banc); Taft v. Taft, 446 N.E.2d 395, 395 (Mass. 1983); Raleigh Fitkin-Paul Morgan Memorial Hosp. v. Anderson, 201 A.2d 537 (N.J.), *cert. denied*, 377 U.S. 983 (1964); Crouse Irving Memorial Hosp., Inc. v. Paddock, 485 N.Y.S.2d 443, 445 (N.Y. Sup. Ct. 1985); *In re* Jamaica Hosp., 491 N.Y.S.2d 898 (N.Y. Sup. Ct. 1985).

23. 410 U.S. 113, 175 (1973).

Roe. In delivery cases, the court may invoke the state's *parens patriae* power to protect children,[24] despite the fact that child protection statutes typically define children as persons upon live birth.[25] At least one court has also recognized the doctors' and hospitals' interests in professional independence and in avoiding risk of liability.[26] In doing so, the court subordinated the woman's interest to those of the fetus.

Most courts seem to perform a simple balancing test in which the state's interest usually prevails.[27] Some of the courts justify forcible intervention by referring to the *Roe v. Wade* trimester analysis.[28] Within the trimester framework, the argument goes, the state's interest in protecting potential life becomes compelling when the fetus becomes viable. The state, therefore, may order the woman to undergo surgery or a transfusion during the third trimester.[29]

An argument being made in legal scholarship and suggested by case law[30] takes this a step further by distinguishing *Roe v. Wade*. According to this argument, if a woman has decided to continue a pregnancy or has failed to terminate the pregnancy within the practical and legal time limits, she waives her right to refuse treatment on behalf of the fetus during the entire pregnancy.[31] A corollary argument would recognize that the fetus has a right to be born sound and healthy.[32]

---

24. *See, e.g., Crouse Irving*, 485 N.Y.S.2d at 445; *Jamaica Hosp.*, 491 N.Y.S.2d at 900 (noting that courts often cite as authority cases in which petitions for medical treatment of children despite parents' objections for religious reasons are granted).

25. *See, e.g., Steven S.*, 178 Cal. Rptr. 525, 527 (1981) ("An unborn fetus is not a person within the meaning of the [child dependency statute]"). *See also In re* A.C., 533 A.2d 611, 616-17 (D.C. App. 1987), *vacated*, 539 A.2d 203 (D.C. App. 1988) (The court in 1987 distinguished between "a court authorizing medical treatment for a child already born and a child who is yet unborn, although the state has compelling interests in protecting the life and health of both children and viable unborn children. Where birth has occurred, the medical treatment does not infringe on the mother's right to bodily integrity. With an unborn child, the state's interest in preserving the health of the child may run squarely against the mother's interest in bodily integrity.").

26. *See, e.g., Crouse Irving*, 485 N.Y.S.2d at 445-46.

27. *See, e.g.*, Jefferson v. Griffin Spaulding County Hosp. Auth., 274 S.E.2d 457, 460 (Ga. 1981).

28. *See, e.g. In re* A.C., 533 A.2d at 614.

29. *See* John Fletcher, *The Fetus as Patient: Ethical Issues*, 246 JAMA 772, 772-73 (1981); Maxwell L. Stearns, *Maternal Duties During Pregnancy: Toward a Conceptual Framework*, 21 NEW ENG. L. REV. 595, 598-606 (1985-86). *See also* Denise K. Cahalane, *Court-Ordered Confinement of Pregnant Women*, 15 CRIM. & CIV. CONFINEMENT 203 (1989). The Court's recent repudiation of the trimester framework in Planned Parenthood v. Casey, 60 U.S.L.W. 4795 (U.S. Jun. 30, 1992) (Nos. 91-744 & 91-902) may make it easier to justify intervention prior to viability.

30. *In re* Jamaica Hosp., 491 N.Y.S.2d 898, 900 (N.Y. Sup. Ct. 1985).

31. John A. Robertson, *The Right to Procreate and in Utero Fetal Therapy*, 3 J. LEGAL MED. 333 (1982). *Contra* SHERMAN ELIAS & GEORGE ANNAS, REPRODUCTIVE GENETICS AND THE LAW 260 (1987).

32. *See, e.g.*, Margery W. Shaw, *Conditional Prospective Rights of the Fetus*,

B. The Moral Formulae

The balancing test in which an interest in protecting fetal life usually outweighs the interests of an individual woman is applied with two types of explanations. One is a call to a greater social good.[33] The other is a lesser of two evils argument.

1. For the Greater Social Good

In this context, a call to the state's interest or "greater social good" asserts both a moral and a quantitative claim. The moral claim speaks of a reverence for the sanctity of life—all forms and stages of human life.[34] Subordinating this state interest, then, risks devaluing the meaning of human life. For many, this claim has religious significance.[35] The quantitative claim is twofold. It suggests that the good of the many members of society who would benefit outweighs the good of the few women subject to forced medical treatment. The good of the many may arise from preserving the sanctity of human life, or from securing a potentially productive next generation, or from both.

From a gendered perspective, the greater social good argument both describes and prescribes the devaluation of women. As a starting point, the argument defines the woman's interests in opposition to society's. The utilitarian argument for the balancing of interests then assumes from the outset that women, or at least pregnant women, are outsiders whose claims are potentially harmful to the concept of ordered liberty. One can point to these cases as proof that women's autonomy interests are not considered to be part of or necessary to

---

5 J. LEGAL MED. 63 (1984). *See also* Deborah Mathieu, *Respecting Liberty and Preventing Harm: Limits of State Interventions in Prenatal Choice*, 8 HARV. J.L. & PUB. POL'Y 19, 49 (1985) (arguing for a woman's limited obligation to prevent serious harm to her future child). However, with Justice Rehnquist's assertion in *Webster v. Reproductive Services* that the state's interest in protecting potential life is compelling before and after viability, the "waiver argument" may not be necessary.

33. *See* Katherine A. Knopoff, Comment, *Can a Pregnant Woman Morally Refuse Fetal Surgery?*, 79 CAL. L. REV. 499, 505-08 (1991) (describing and rejecting utilitarianism as a possible basis for considering the morality of forced fetal surgery).

34. For other expression of this claim see Webster v. Reproductive Health Servs., 490 U.S. 492, 501, 504-07 (1989); John T. Noonan, Jr., *An Almost Absolute Value in History, in* THE MORALITY OF ABORTION: LEGAL AND HISTORICAL PERSPECTIVES 51-59 (John T. Noonan ed. 1970).

35. *See, e.g.*, Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130, 1132 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986) (In seeking order to have patient Ernestine Jackson submit to blood transfusion, Mercy Hospital, run by the Catholic religious order "dedicated to the preservation of life and family," argued that denial of a guardianship would uphold Mrs. Jackson's religious beliefs to the detriment of the hospital's own religious convictions.).

the greater social good, and that the social good is defined by reference to the interests of men.

As a prescription for social order, the balancing test suggests a hierarchy of interests, with those of women subordinated to those of the "greater good." Indeed, the constitutional formula—the trimester framework—expressly presumes that the woman's interests will be subordinated at some inevitable point in time—at least upon fetal viability, if not earlier. The presumption that the woman's interests can and will be subordinated describes her as a means of perpetuating hierarchical order. The fact that the balancing test is applied vis-a-vis woman's biological capacity to bear children exacerbates this reductionist view of women. Thus, the law describes women first and foremost as childbearers. This role-specific definition of women, in turn, limits their ability to assert a broader and more heavily weighted set of interests. The descriptive and prescriptive effect of the balancing test comes full circle.[36]

2. The Lesser of Two Evils

The lesser of two evils argument focuses on the potential harm to the woman's interests and the fetus', and it assumes those harms can be quantified. Where the woman's interests are quantifiably less, the court will order her to submit to medical treatment. This argument is more closely tied to the facts of any case. The focus is more clearly on the particular woman, and the fetus is both particular and representative of the greater social good. This argument has been made in two different ways.

One way is to point to the assessment of medical risk to woman and fetus and to weigh both of those risks against the woman's

---

36. If one assumes that women's rights have not been devalued relative to men's, then fetuses take on a superhuman quality, able to trump the rights of women and men with a single claim. This creates two negative implications for women. In the forced medical treatment cases, the woman's refusal to consent to therapy justifies state intervention on behalf of the fetus. The fact that the woman carries the fetus makes it more likely that her choices will be restricted. Thus women are still being regulated because of their biological capacity to bear children and are therefore still being described as childbearers. In addition, if the assumption of equality is correct, it seems possible to imagine a court ordering a man to donate blood, bone marrow, or other tissue necessary to fetal health. Hart v. Brown, 289 A.2d 386 (Conn. 1972) (court authorized parents to consent to kidney transplant between identical twin, aged 7 years and 10 months); Strunk v. Strunk, 445 S.W.2d 145 (Ky. 1969) (court authorized kidney donation from mentally incompetent brother to brother dying of kidney disease). *But see* McFall v. Shimp, 10 Pa. D. & C.3d 90 (C.C.P. Allegheny Co. Pa. 1978) (court has no authority to compel a relative to participate in bone marrow transplant). Or perhaps a more likely scenario would have a court ordering a man to refrain from fetus-risking activity. This would obviously substract from the rights held by all adults, both women and men.

autonomy interests. For example, the Georgia Supreme Court denied Jessie Mae Jefferson's motion for stay of a court-ordered cesarean where the attending physician stated that because of a complete placenta previa, there was "99% certainty that the child cannot survive natural childbirth," that "[t]he chances of defendant surviving vaginal delivery are no better than 50%," and that "a delivery by caesarean section . . . would have an almost 100% chance of preserving the life of the child, along with that of defendant."[37] After reciting these statements, the court affirmed the trial court's conclusion that "the intrusion involved into the life of Jessie Mae Jefferson and her husband, John W. Jefferson, is outweighed by the duty of the State to protect a living, unborn human being from meeting his or her death before being given an opportunity to live."[38]

If the *Jefferson* opinion accurately describes the court's analysis, the court weighed the assessed medical risk to woman and fetus against the assessed medical benefit to the fetus and the unspecified weight of the woman's interests. It appears that the court weighed apples (medical risks) against an orange (medical benefit) and a banana (woman's interests in bodily integrity and self-determination). Therefore, one could conclude, using algebraic logic, that the woman's interests weighed relatively little. In addition, by balancing the assessed medical risk to the woman against her autonomy interests, the court contradicts the notion that the autonomy interests are the woman's and not the state's.

The Georgia court's conclusion also illustrates the second way to argue that forced medical intervention is the lesser of two evils—to assume harm to the woman's interests is limited in time. This argument admits that ordering treatment interferes with the woman's autonomy interests and that cesarean surgery poses risk to the woman's physical condition. According to this argument, however, the effects of performing a cesarean or other treatment are temporary, or at least, short-lived compared to the alternative. Judge Nebeker's opinion for the District of Columbia Court of Appeals in *In re A.C.*[39] presents this argument in its extreme form. The court denied

---

37. Jefferson v. Griffin Spaulding County Hosp. Auth., 274 S.E.2d 457, 458 (Ga. 1981).

38. *Id.* at 460. *See also Mercy Hosp.*, 489 A.2d at 1130 (Maryland appellate court affirmed lower court's decision to deny hospital's petition to appoint guardian for competent woman where the likelihood that woman would need transfusion during cesarean was 40%-50% and there was no risk to baby); Taft v. Taft, 446 N.E.2d 395, 397 (Mass. 1983) (Massachusetts Supreme Court vacated lower court's order of forced cerclage operation to hold pregnancy for lack of evidence. "We have no findings, based on expert testimony, describing the operative procedure, stating the nature of any risks to the wife and to the unborn child . . . . We have no showing of the degree of likelihood that the pregnancy will be carried to term without the operation.").

39. 533 A.2d 611 (1987), *vacated*, 539 A.2d 203 (D.C. 1988).

a motion for stay of a cesarean section on a terminally ill woman, Angela Carder, whose consent to surgery was equivocal. In finding that the "trial judge did not err in subordinating A.C.'s right against bodily intrusion to the interests of the unborn child and the state," the court stated that "[t]he Caesarean section would not significantly affect A.C.'s condition because she had, at best, two days left of sedated life."[40] Thus the court put Angela Carder's interests on a sliding life expectancy scale.

The lesser of two evils argument assumes that autonomy interests can and should be quantified, and in the balancing process, the courts assign little weight to the woman's interests. As a description of the existing social order, the argument reinforces the point made above that recognizing a state interest in protecting fetuses is yet another justification for trumping women's rights. And the description in turn furthers the imbalance in the weighting of interests. If one assumes that the interest in protecting fetal life weighs heavily relative to the woman's interests, then one can more easily ignore Jessie Mae Jefferson's claim that as a person she is entitled to have her interests accorded great respect,[41] that as a competent adult she, and not the state, should weigh the assessed medical risks against her autonomy interests,[42] and that she should not be compelled to undergo risk against her will because of biological capacity to bear children.[43] If one assumes that pregnancy diminishes the woman's interests, then one can more easily deny that Angela Carder's liberty interests might increase as the significance of the decision to her own life increases.

C. Conclusion of the Standard Story

In nearly all of the cases reported in medical and legal journals, it was the doctors and/or hospital who petitioned the court. The outcome, as the 1987 survey indicates, strongly favors the medical factors over the woman's personal decision. Curiously, the standard story begins inappositely, by recognizing the individual's interests. It then progresses to a balancing of interests. This formulaic approach,

---

40. *Id.* at 617.

41. *See* Dawn E. Johnsen, *The Creation of Fetal Rights: Conflicts with Women's Constitutional Rights to Liberty, Privacy, and Equal Protection*, 95 YALE L.J. 599 (1986); Lawrence J. Nelson et al., *Forced Medical Treatment of Pregnant Women: "Compelling Each to Live as Seems Good to the Rest,"* 37 HASTINGS L.J. 703, 750 (1986).

42. *See* Janet Gallagher, *Prenatal Invasions & Interventions: What's Wrong With Fetal Rights*, 10 HARV. WOMEN'S L.J. 9, 14 (1987).

43. *See* Nancy K. Rhoden, *The Judge in the Delivery Room: The Emergence of Court-Ordered Cesareans*, 74 CAL. L. REV. 1951 (1986) [hereinafter Rhoden, *Delivery Room*].

according to traditional legal assumptions, prescribes consistent, even-handed, and neutral decisionmaking.[44] The standard story's integrity stands, if one does not question these conclusions. Gendered questioning, however, reveals that the story really begins by devaluing the woman's interests; it then progresses to a weighted balancing premised on a concept of social good and ordered liberty that excludes women.

II. The Response With Respect to All Women

The critique of the standard story may suggest several responses.[45] But so far, the response has been fairly uniform—a call to reclaim rights against the state.[46] Part II describes the call to rights with respect to its effect on women's status and then questions whether it is merely a result dictated by the standard story.

A. Reclaiming Rights

This call to rights comes as a reminder that the principles of bodily integrity and decisional autonomy form a line against state interference and that these rights are so central to our political and social understanding of liberty that they cannot be abrogated without restricting our understanding of freedom.[47]

Those who make the call to rights bolster it with several types of arguments. Two of these arguments speak to the weighting of interests. One of these recasts the balancing test as a comparison between the woman's constitutional status and the fetus'. The woman's "personhood is free from doubt," while the fetus' "legal status as a person is uncertain."[48] This view challenges the weighting of fetal interests as heavier than the woman's. The second follows from the *Roe v. Wade* balancing test, which recognizes a compelling state interest in protecting maternal health.[49] Because the state cannot prohibit an abortion necessary to save the life or health of the woman, it cannot force her "to undergo medical treatment for the sake of the fetus if that treatment endangers her life or health in

---

44. *But see* Robert F. Nagel, *The Formulaic Constitution*, 84 Mich. L. Rev. 165 (1985).

45. For a more general discussion of feminist responses to gender discrimination, see Deborah L. Rhode, *The "No-Problem" Problem: Feminist Challenges and Cultural Change*, 100 Yale L.J. 1731 (1991).

46. *But see* Patricia Williams, *Fetal Fictions: An Exploration of Property Archetypes in Racial and Gendered Contexts*, 42 Fla. L. Rev. 81 (1990).

47. *See, e.g.*, Gallagher, *supra* note 42, at 57; Johnsen, *supra* note 41, at 614-15; Nelson *supra* note 41, at 750.

48. Nelson, *supra* note 41, at 749.

49. Roe v. Wade, 410 U.S. 113, 164-65 (1973).

any way."[50] Because most medical procedures create some risk, the state interest in protecting maternal health is present to some degree in most cases.

Two other arguments reject the balancing test and address the construct of the woman-fetus relationship. One of these points to the good samaritan doctrine, which prohibits the state from compelling one person to aid or rescue another.[51] The doctrine, based on principles of bodily integrity and decisional autonomy, arguably applies to pregnancy. The state cannot compel a woman to protect the fetus without substituting paternalism for choice, thus compromising the integrity of choice[52] as a liberal goal and negating the moral force of altrusim.[53] The other argument redescribes the relevant relationship as that between the fetus and society, with the woman and not the court as the most appropriate mediator.[54] This gives the individual woman the authority to weigh the interests. It also acknowledges the intimate nature of the decision.

Yet another set of arguments focuses on the effects of court-ordered medical treatment of pregnant women. One such argument expresses abhorrence at a direct result of court orders—use of state-sanctioned force against individuals.[55] Once the court grants a petition, the state becomes morally responsible for using physical, sometimes violent, force against an individual.[56] Even if physical violence is not necessary, the order mentally coerces the woman.[57] And the use of coercion, whether physical or mental, also proscribes our understanding and expectation of liberty. The second of these arguments predicts that this use of force will discourage women from seeking medical care in the first place, thus increasing the risk of harm to pregnant women and fetuses.[58]

---

50. *See, e.g.*, Nelson, *supra* note 41, at 749-750. *See also* Susan Goldberg, *Medical Choices During Pregnancy: Whose Decision Is It Anyway?*, 41 RUTGERS L. REV. 591, 599 (1989).

51. W. PAGE KEETON, ET AL., PROSSER & KEETON ON THE LAW OF TORTS § 56, at 375 (5th ed. 1984).

52. Rhoden, *Delivery Room*, *supra* note 43, at 2005-06, 2029-30.

53. *Id.* at 1982.

54. *See* Gallagher, *supra* note 42, at 13.

55. *See* George Annas, *Foreclosing the Use of Force: A.C. Reversed*, 20 HASTINGS CENTER REP. 27 (July/Aug. 1990); Rhoden, *Delivery Room*, *supra* note 43, at 2003.

56. At least one woman has been literally tied down. Seven security officers forcibly removed her husband from the hospital. Gallagher, *supra* note 42, at 9-10; Kolder, *supra* note 9 at 1193. *See also infra* note 141 and accompanying text.

57. *See, e.g.*, Bowes & Selgestad, *supra* note 21, at 210 ("Following the court's decision, the patient, although still reluctant, became more cooperative and agreed to the inducement of general anesthesia."); Kolder, *supra* note 9, at 1193 ("Two of 13 court orders [for cesarean sections] were not enforced because the patient finally agreed to the procedure.").

58. *See* Goldberg, *supra* note 50, at 620.

The last argument used to bolster the call to rights is the slippery slope argument.[59] If the state may order forced blood transfusions and cesareans, it may also justify fetal surgeries without the woman's consent. Further, in order to protect the state's interest in healthy fetal life, it may compel a woman to take medication, even vitamins, to seek prenatal care, to avoid intake of alcohol, drugs and teratogens, and other risk-creating activity. This argument is significant in that it is frequently made, it is often used to bolster the other arguments, and because it expresses a last stand mentality. The argument assumes that once started, the additional subtraction of women's rights will not be checked.

B. The Standard Story Retold

The odd thing about the response to the forced medical treatment of pregnant women is that it has generated such a uniform response—the call to rights. In the context of an issue that so clearly illustrates the subordination of women, and given the richness of feminist and critical theory and method, that singular response seems very unresponsive. True, critics of the standard story offer a variety of arguments to support the call to rights. But they have not invoked any of the alternatives to traditional rights analysis that feminists and critical theorists have developed in other contexts.[60] It seems that those who have responded critically to the standard legal story have not abandoned liberalism, but reject it in its current gender-biased form.

What these critics have also done is to respond directly and solely to the standard story. Thus, they have allowed the standard story to dictate the boundaries of the inquiry. The gendered inquiry into the forced medical treatment of pregnant women begins much as the standard story does, by referring to the woman's right to bodily integrity and decisional autonomy. The inquiry then challenges the legal and moral formulae but does so only in terms of maternal and fetal interests, thereby accepting that those who tell the standard story may define the parameters for discussing subordination.

The call to rights has made clear that these cases are not simply about pregnancy and potential life; that they could not occur but

---

59. *See, e.g.*, Nelson, *supra* note 41, at 756-757; Rhoden, *Delivery Room*, *supra* note 43, at 2023-29. *See also* Knopoff, *supra* note 33 (anticipating and arguing against forcible fetal surgery); Robertson, *supra* note 31 (anticipating and supporting forcible fetal surgery during the third trimester in states that have prohibited post-viability abortions); Shaw, *supra* note 31, at 66 (proposing a duty to undergo a broad array of prenatal treatments "if it is expected to benefit the would-be child").

60. For example, reformulating rights to address social hierarchy, implementing feminist methods such as consciousness raising, and incorporating values identified with the female experience into the definition of the social good.

for the prior devaluation of women; that "fetal interests" is a proxy for majoritarian interests; and that the utilitarian balancing test describes women as tools useful for serving the rest of society. But few, if any, have acknowledged that these cases prove the call to rights alone has already failed. Few, if any, also have acknowledged that this failure may arise from the lack of a full understanding of how patriarchy is perpetuated. More specifically, this failure may arise in the first instance from a failure of inquiry. If we limit the inquiry to the standard story, we can only respond in those terms.

Therefore, I argue that the standard story not only masks gender-biased reasons for ordering medical treatment against the woman's will, but it also limits the inquiry. This is no deliberate, malevolent conspiracy. It occurs in part because traditional legal discourse posits both judicial neutrality and normative rule-making as goals of our legal system. Judges must then appear neutral while imposing normative moral judgments. The mechanism is the formula—here, the balancing test. Scales empty, the balancing test is neutral. It is also meaningless. With scales loaded, the balance tips, and because we know the scale itself is value neutral, we assume the weighing merely reflects the physical laws of pregnancy. The formula masks the fact that value-based decisionmaking selects the interests and assigns them weight.[61] The call to rights response challenges the weighting of the interests, but it fails to reveal how the woman's interests came to weigh so little in the first place.

The standard story also effects a limited inquiry by directing attention to the moral force of "fetal interests." This makes it politically difficult and socially unacceptable to deny that preventing harm to a fetus is good or that fetal injury or death is tragic. In fact, there is no reason to issue a denial. We need not step all over ourselves in order to avoid appearing callous to future babies or apologetic in our defense of women's interests. We need not seek compromises that would allow fetal interests to prevail some of the time but not as often as current cases permit.[62] We can agree that potential life is important without conceding that protecting fetal interests is the primary and central issue. We need to seize the center. Looking to the effect of the standard story on women's lives and looking to the stories told by women is one way to do so.[63]

---

61. *See* Richard Delgado, *Judicial Influences and the Inside-Outside Dichotomy: A Comment on Professor Nagel*, 61 U. COLO. L. REV. 711 (1990); Mari J. Matsuda, *Liberal Jurisprudence and Abstracted Visions of Human Nature: A Feminist Critique of Rawls' Theory of Justice*, 16 N.M. L. REV. 613 (1986); Nagel, *supra* note 44.

62. *See, e.g.*, Gallagher, *supra* note 42, at 54-56; Deborah Mathieu, *Respecting Liberty and Preventing Harm: Limits of State Intervention in Prenatal Choice*, 8 HARV. J.L. & PUB. POL'Y 19, 45-54

63. Matsuda, *Pragmatism Modified*, *supra* note 5, at 1764-68; Richard Del-

III. THE SOCIAL CONSTRUCTION OF GENDER IN THE MATERNAL-FETAL CONFLICT

So far we have examined two stories about pregnant women. The first describes women as childbearers. It focuses on the biological capacity to reproduce. From that, it extracts a moral duty to make decisions on behalf of the fetus without regard to self but perhaps with regard to a greater or higher good. When that greater good becomes synonymous with state interest, it can be enforced at law. Another version rejects the notion that women can be regulated as childbearers. It criticizes the effect of women as such as essentializing, reducing them to biological vessels whose interests can be justifiably subordinated. It reclaims or redefines rights and depicts a liberalism that is not gender-biased.

The response to the standard story, however, is premature. One cannot call for an end to gender bias without understanding how gender is constructed. Further inquiry is necessary. The cases must be questioned again. We must look beyond the standard legal story to the descriptions of pregnant women, mothers, and doctors. We must look to the gaps between the standard story and the social, political, and economic reality in which women live. This inquiry can begin with the social construction of gender (Part III) but also delves further into the effects of racism, classism, and cultural imperialism on the lives of women (Part IV) so that we may gain a fuller understanding of the nature of patriarchy and respond to it in ways that reach each woman.

A. Stories About Pregnant Women

Looking beyond the formulae means looking to a time before the invocation of rights and the balancing of interests. It means looking beyond the law, to the way that social reality constructs the conflict. The apparent conflict which takes place between the woman and the doctor provides a starting point.

1. Judges Listen to Doctors

The contents of judicial opinions reflect the relative significance accorded to doctors' and women's views. For example, the *per curiam* opinion in *Jefferson v. Griffin Spalding County Hospital*[64] "readopts" the findings made by the superior court that initially granted the hospital's petition for custody of the unborn child. The opinion

---

gado, *Storytelling for Oppositionists and Others: A Plea for Narrative*, 87 MICH. L. REV. 2411, 2435-41 (1989) [hereinafter Delgado, *Storytelling*].

64. 274 S.E.2d 457 (Ga. 1981).

then restates the doctors' assessment of risk in one long paragraph.[65] In the paragraph's last sentence, the Court finds "that as a matter of fact the child is a human being fully capable of sustaining life independent of the mother."[66] This enabled the court to grant temporary custody of the unborn child to the state.

The opinion summarizes the Jeffersons' position in three sentences:

> Mrs. Jefferson and her husband have refused and continue to refuse to give consent to a Caesarean section. This refusal is based entirely on the religious beliefs of Mr. and Mrs. Jefferson. They are of the view that the Lord has healed her body and that whatever happens to the child will be the Lord's will.[67]

The order to submit to a Caesarean section was conditioned on the results of a sonagram. The sonagram, when performed, showed that Mrs. Jefferson was no longer placenta previa.[68] She delivered a healthy boy by vaginal birth.[69]

The opinion in *In re Jamaica Hospital*[70] also emphasizes the importance of the medical view relative to the woman's. Judge Lonschein mentions the hospital attorney's first and last name and his law firm. He gives us the attending physicians' full names. He refers to the woman whose life and pregnancy are in danger as "the patient." This may have been done to protect her privacy. But the limited time he spent eliciting her view and the minimal space he used to present it indicate that he excluded from consideration other aspects of her identity as well. Judge Lonschein dedicated much of three long paragraphs to describing the doctor's opinion. He repeated each of the medical conclusions—that the woman was eighteen weeks pregnant, that she was bleeding, that her hemoglobin and hematocrit readings were below normal, that she may die without a transfusion and that the fetus would die without a transfusion—at least twice.[71] Of the woman's position, he tells us that she is a Jehovah's Witness and that when asked if she would consent to a blood transfusion,

---

65. Based on the evidence presented, the Court finds that Jessie Mae Jefferson is due to begin labor at any moment. There is a 99 to 100 percent certainty that the unborn child will die if she attempts to have the child by vaginal delivery. There is a 99 to 100 percent chance that the child will live if the baby is delivered by Caesarean section prior to the beginning of labor. There is a 50 percent chance that Mrs. Jefferson herself will die if vaginal delivery is attempted. There is an almost 100 percent chance that Mrs. Jefferson will survive if a delivery by Caesarean section is done prior to the beginning of labor.
Id. at 459.

66. *Id.*

67. *Id.*

68. *Id.* at 461 n.1.

69. *See infra* notes 128-36 and accompanying text.

70. 491 N.Y.S.2d 898 (N.Y. Sup. Ct. 1985).

71. *Id.* at 899.

"[s]he told me, in effect, that because of her religion, she would not."[72] The judge then dismisses the woman's privacy and first amendment interests in very short order and decides that the 18-week-old fetus "can be regarded as a human being, to whom the court stands in parens patriae." He appoints the attending physician as special guardian of the fetus.[73]

One might explain the differences in time and space spent on the medical and woman's views in gender-neutral terms. The space allocation may reflect the simple fact that the doctors' views were more detailed and thus, that more space was necessary to accurately describe those views. The space allocations in the opinion may reflect the conclusion the judge reached only after carefully and evenly balancing the two views. On the other hand, these opinions may illustrate that socially constructed biases are at work—judges listen to doctors because they are usually also male professionals and therefore presumptively rational; judges listen to doctors because medicine is regarded as a source of authority whereas individual women are regarded as a source of trouble. In other words, the outcome of these cases may be partially explained by the fact that stereotypes about men, doctors, women, and pregnant women inform the law.[74]

2. Doctors and Judges Discount the Woman's Voice

At the same time that judges attach greater relative weight to the doctor's opinion, both doctors and judges discount the woman's refusal to consent to treatment. When women refuse they are often characterized as stubborn, guilty, and irrational, even when the court specifically finds them to be clearly competent. One case report described the patient as "angry and uncooperative."[75] She refused to consent to a cesarean delivery "[b]ecause of her fear of surgery." Yet, the medical staff "viewed the patient's response as one of unreasonable insensitivity to the welfare of her infant."[76] The report describes their "devastating sense of helplessness" but not the patient's. Another report by four doctors describes two cases. In Case 1, "[c]esarean section was proposed to save the life of the fetus. The patient, however, stubbornly refused to submit to surgery."[77] The report speculates that her refusal expressed guilt over an unplanned

---

72. *Id.*
73. *Id.* at 900.
74. *See infra*, Part IV.A.
75. Bowes & Selgestad, *supra* note 21, at 210.
76. *Id.* at 209.
77. J.R. Leiberman et al., *The Fetal Right to Live*, 53 OBSTET. & GYN. 515, 515 (1979).

pregnancy.[78] In Case 2, the fetus was stillborn after the woman refused to consent to cesarean surgery because she was afraid of dying.[79] The discussion following the case states "it is logical to regard it as a real felony."[80]

Even holding to a religious belief may reflect weakness, not strength. In *Crouse Irving Memorial Hospital, Inc. v. Paddock*,[81] Mrs. Paddock consented to a cesarean section but refused blood transfusions for religious reasons. The opinion said, "Mrs. Paddock is an adult obviously of sound mind and deep religious conviction."[82] Then it described her consent as contradictory and irrational:

> She wants the hospital and her doctors to take aggressive medical steps to insure a proper delivery, but does not want the medical personnel to correct a possible grave condition which may unavoidably be encountered in the process. This, it seems to me, puts the hospital and her doctors in an untenable position.[83]

The tone of the opinion suggests the court believed Mrs. Paddock had victimized the hospital, depriving it of freedom and integrity: "A hospital is not the patient's servant, subject to his orders. The hospital shares the physician's independence of judgment and responsibility for action, and to let a patient die runs counter to the reasons for the hospital's existence."[84]

In the reported cases, only one trial judge refused to issue an order enabling forced medical treatment,[85] three appellate judges vacated lower court orders,[86] and one appellate court vacated an order on rehearing.[87] In three of the appellate cases, however, the trial court's orders had been carried out before the appeal.[88] Only Ernestine Jackson and Susan Taft were not compelled to submit to medical treatment they had refused.[89] Even so, they were compelled

---

78. *Id.* at 516.
79. *Id.*
80. *Id.* at 517.
81. 485 N.Y.S.2d 443 (N.Y. Sup. Ct. 1985).
82. *Id.* at 445.
83. *Id.*
84. *Id.* (quoting David J. Sharpe & Robert F. Hargest, III, *Lifesaving Treatment for Unwilling Patients*, 36 FORDHAM L. REV. 695, 701 (1968)).
85. Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130, 1134 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986).
86. *In re* Steven S., 178 Cal. Rptr. 525, 529 (Cal. Ct. App. 1981); Taft v. Taft, 446 N.E.2d 395, 397 (Mass. 1983); Fosmire v. Nicoleau, 551 N.E.2d 77, 84 (N.Y. App. Div. 1990), *aff'g* 536 N.Y.S.2d 592 (N.Y. App. Div. 1989).
87. *In re* A.C., 573 A.2d 1235, 1253 (D.C. 1990) (en banc).
88. *See Id.* at 1241; *Steven S.*, 178 Cal. Rptr. at 526; *Fosmire*, 551 N.E.2d at 79.
89. Jessie Mae Jefferson was ordered to undergo a sonagram and a cesarean. She was compelled to undergo a sonogram, but the results of that test proved the cesarean was unnecessary. Jefferson v. Griffin Spalding County Hosp. Auth., 274 S.E.2d 457, 460 (Ga. 1981).

to participate in two hearings each in order to preserve their right to bodily integrity and decisional autonomy. Ernestine Jackson and Susan Taft were also the only two women expressly described as competent, conscious and rational adults.[90] The gender-neutral explanation—that only Ernestine Jackson and Susan Taft were competent—does not work here. It is clear in all but two cases[91] that the women were conscious, rational, and competent to decide. The more likely explanation is that courts are reluctant to assume that a woman, especially a pregnant woman, is rational, and competent; they may be, in fact, applying the stereotype—that pregnant women are subject to the whims of their ever-changing hormonal imbalance and are incapable of knowing their own minds, or that they are morally culpable.

B. Stories About Doctors and Men

Doctors and other medical staff, on the other hand, are presumptively rational, steady, and well-motivated. It is recognized that by having two patients—mother and fetus, doctors have a difficult task. And it is assumed that they are motivated primarily by their ethical duty to care for both patients.

The point of view from which these cases are discussed strikingly illustrates the presumption in favor of doctors. The reports not only allocate more time and space to the medical opinion, but they introduce and describe the woman in the physician's terms. "At the time she was 36 years old and, except for the loss of blood, apparently in good health."[92] "Stacey Paddock is pregnant. She has what is known as an intrauterine pregnancy which is further complicated by the fact that she is anemic, by her Rh negative blood type and by the anterior position of her placenta. Furthermore, according to her attending physician, Dr. M. Robert Neulander, her blood count is already low."[93] "Defendant is in the thirty-ninth week of pregnancy."[94] The medical facts are legally relevant, but the reports indicate that the judge saw the woman first and foremost as the doctor's patient.

When the doctor or hospital petitions for an order, and the judge issues an order for forced medical treatment, it is always because it

---

90. Note that two other reports make specific findings of competency but then qualify that finding by characterizing the women as unreasonable. Crouse Irving Memorial Hosp., Inc. v. Paddock, 485 N.Y.S.2d 443, 445 (N.Y. Sup. Ct. 1985); Bowes & Selgestad, *supra* note 21, at 209.

91. *See A.C.*, 573 A.2d at 1235; *In re* Steven S., 178 Cal. Rptr. 525 (Cal. Ct. App. 1981).

92. Fosmire v. Nicoleau, 551 N.E.2d 72, 77 (N.Y. App. Div. 1990).

93. *Crouse Irving*, 485 N.Y.S.2d at 444.

94. Jefferson v. Griffin Spalding County Hosp. Auth., 274 S.E.2d 457, 458 (Ga. 1981).

is "necessary to save the life" of the fetus, of the woman, or both.[95] Some reports depict a fetus imperiled by her mother's condition, casting the doctor and judge into the role of rescuing heroes. Judge Lonschein's opinion in *In re Jamaica Hospital* again clearly illustrates the point. He wrote the opinion in first person, and his initial sentences describe his own activities. "This past Saturday evening, April 20th, 1985, at about 6 p.m., while I was getting dressed for a dinner engagement, I received a telephone call at my home . . . . "[96] Thus he is called to action on behalf of the fetus "in mortal danger,[97] . . . in a life-threatening situation."[98] He responds to "the danger to the fetus,"[99] of course, by issuing the order. "I therefore appointed Dr. Capiello as special guardian of the unborn child and ordered him to exercise his discretion to do all that in his medical judgment was necessary to save its life."[100] The use of first person apparently indicates that Judge Lonschein personally took responsibility for the rescue.[101]

This casting echoes the medical perspective on childbirth, which centers on the physician's role in "the management of labor."[102] The understanding of childbirth as a pathology, a set of risks to be controlled, developed during the twentieth century as doctors became participants in caring for pregnant women.[103] Medical intervention has steadily increased through the past few decades. Cesarean sections, in particular, have risen in number[104] although the mortality

---

95. *See, e.g.*, *id.* 460; Raleigh Fitkin-Paul Morgan Memorial Hosp. v. Anderson, 201 A.2d 537, 538 (N.J.), *cert. denied*, 377 U.S. 983 (1964); Crouse Irving Memorial Hosp., Inc. v. Paddock, 485 N.Y.S.2d 443, 444-45 (N.Y. Sup. Ct. 1985); *Jamaica Hosp.*, 491 N.Y.S.2d at 900; *Fosmire*, 551 N.E.2d at 79.

96. *In re* Jamaica Hosp., 491 N.Y.S.2d 898, 898-99 (N.Y. Sup. Ct. 1985).

97. *Id.* at 899.

98. *Id.*

99. *Id.* at 900. *See also Raleigh Fitkin-Paul*, 201 A.2d at 538.

100. *Jamaica Hosp.*, 491 N.Y.S.2d at 900.

101. *See* Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130, 1132 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986). Chief Judge Gilbert also described his role as central to the drama, but did so by cloaking himself in his legal duties. "The State, in this case personified by the circuit court judge, must 'be a neutral in its relations with groups of religious believers and non-believers'." *Id.*

102. Bowes & Selgestad, *supra* note 21, at 211.

103. JUDITH W. LEAVITT, BROUGHT TO BED: CHILDBEARING IN AMERICA: 1750 TO 1950, 38-40, 140, 174-179 (1986); RICHARD W. WERTZ & DOROTHY C. WERTZ, LYING-IN: A HISTORY OF CHILDBIRTH IN AMERICA 133 (1977). *See also* JOHN S. HALLER & ROBIN N. HALLER, THE PHYSICIAN AND SEXUALITY IN VICTORIAN AMERICA x-xii (1974).

104. Gertrud S. Berkowitz et al., *Effect of Physician Characteristics on the Cesarean Birth Rate*, 161 OBSTET. & GYN. 146 (1989) (between 1970 and 1985, the cesarean birth rate in the United States has increased from 5.5% to 22.7%); Helen I. Marieskind, *Cesarean Section in the United States: Has It Changed Since 1979?*,

rate for women undergoing cesarean surgery is significantly higher than the mortality rate for women who deliver vaginally.[105] As a corollary to the medicalization of pregnancy and childbirth, women have been displaced as actors in the reproductive process; they have become sources of risk and conflict.

Finally, note that in the few cases in which a court has refused to order forced medical treatment or where an appellate court vacated the order, the woman was married. In two of the four reported cases, the husband clearly supported the woman's decision.[106] In *Taft v. Taft*,[107] the husband initiated the proceeding to have Susan Taft undergo cerclage surgery, despite her religious objections. The Massachusetts Supreme Court vacated the judgment for lack of evidence regarding the medical necessity for the operation, but did not decide whether the husband had standing.[108] And in *In re A.C.*, Angela Carder's husband was "too distraught to testify and uttered only a few words at the hearing,"[109] but other members of her family and her doctor opposed cesarean surgery on her behalf.[110] So, it seems that while judges may discount the woman's refusal to consent to treatment, they may listen to others.

The implications are particularly frightening for single women. According to the 1987 national survey, forty-four percent of the women in cases where court orders were sought were unmarried.[111] The published cases suggest that a court probably would honor only an unmarried woman's right to refuse medical treatment if her doctor

---

16 BIRTH 196, 196 (1989) ("In 1979, the cesarean section rate in the United States was 16.4 percent, having risen 264 percent since 1965, when the National Center of Health Statistics recorded a rate of 4.5 percent. Several factors were identified as contributing to the rise . . . . By 1987, the most recent year for which date are available, the rate was 24.4 percent nationally—an increase of 48.8 percent in the 8 years since 1979 and of 442.2 percent in the 22 years since 1965."); Richard P. Porreco et al., *Commentaries: The Cesarean Section Rate is 25 Percent and Rising: Why? What Can Be Done About It?*, 16 BIRTH 118, 119 (1989) (citing cesarean rates of 25 to 35%).

105. *See In re* A.C., 533 A.2d 611, 617, n.5 (1987), *vacated*, 539 A.2d 203 (D.C. App. 1988). (The prior court stated "the death rate of women upon whom Caesarean sections have been performed is between 0.1 percent and 1 percent, significantly higher than the death rate of women who have delivered their babies vaginally.").

106. Mercy Hosp., Inc. v. Jackson, 489 A.2d 1130, 1131 (Md. Ct. Spec. App. 1985), *vacated*, 510 A.2d 562 (Ct. App. 1986); Fosmire v. Nicoleau, 551 N.E.2d 77, 78 (N.Y. App. Div. 1990).

107. 446 N.E.2d 395 (Mass. 1983).

108. *Id.* at 396 ("We shall assume, for the purposes of this case only, that the question of requiring the operation was properly before the judge for his decision.").

109. *In re* A.C., 573 A.2d at 1240 n.4.

110. *Id.* at 1239.

111. Kolder, *supra* note 9, at 1193.

concurred. Because the doctor initiates the proceeding in most cases, it appears very unlikely that a court would deny a petition to have a single woman submit to surgery, detention, or transfusion.

It is as if the woman is the only person who lacks standing in these cases. More accurately, the social devoicing of women both precedes and controls application of the legal formulae, making her legal standing irrelevant. The formal legal analysis begins by presuming an autonomous, independent individual. This presumption would bestow an equal range of choices for men and women. But the fact of social hierarchy deprives women of choice before it is offered. And when choice is formally offered, it occurs as a weighing of preconstructed interests by a judge living within a reality where the privileging of men and doctors and the devoicing of women is no longer obvious. The standard story only expresses pre-existing social conclusions.[112] And a response tailored to (and by) the legal formula only perpetuates the denial of status that occurs within our private political lives.

C. Possible Responses

Challenging the formulae in solely formulaic terms will not address the gender hierarchy described in these cases, the one embedded in our day-to-day lives. Any legal theory that addresses subordination must describe and prescribe a transformation of local or private politics. The study of the social construction of reality enables this understanding, and critical feminists have articulated a variety of transformative theories and methods.[113] It is not the purpose of this article to describe the richness of critical feminist scholarship and

---

112. The privileging of men and doctors and the devoicing of women has becomes apparent in other medical decisionmaking contexts, as well. *See, e.g.*, Steven H. Miles & Allison August, *Courts, Gender and "The Right to Die,"* 18 LAW MED. & HEALTH CARE 85, 85, 87 (1989) "Judicial reasons about profoundly ill, incompetent men accepts evidence of mens' treatment preferences . . . Judicial reasoning about women defines the role of caregivers in making treatment decisions after either rejecting or failing to consider evidence of women's preferences." The authors examined appellate court rulings and found four major gender differences in how courts weigh the previously competent individual's moral preferences:

> The first difference is the courts' view that a man's opinions are rational and a woman's remarks are unreflective, emotional, or immature. Second, women's moral agency in relation to medical decisions is often not recognized. Third, courts apply evidentiary standards differently to evidence about men's and women's preferences. Fourth, life-support dependent men are seen as subjected to medical assault; women are seen as vulnerable to medical neglect.

113. *See, e.g.*, CATHARINE A. MACKINNON, FEMINISM UNMODIFIED (1987); Lucinda M. Findley, *Transcending Equality Theory: A Way Out of Maternity and the Workplace Debate*, 86 COLUM. L. REV. 1118 (1986); Deborah L. Rhode, *Feminist Critical Theories*, 42 STAN. L. REV. 617 (1990).

activity but rather to point to its possibilities while recognizing its limitations.

Feminist scholars like Catharine MacKinnon and Robin West have suggested an understanding of rights that goes beyond the negative right of constitutional theory.[114] They have also proposed framing harms in terms "true to our own experience and our own subjective lives."[115] And they have proposed these new formulations as transformative tools.

The concept of negative rights is ineffective where subordination is not simply the result of government oppression but is the social order. This reflects the understanding that "[b]ecause political power in [the modern state] could emancipate the individual only within the framework of the existing social order, law could emancipate women to be equal only within 'the slavery of civil society.'"[116] A right against discrimination by the State does not begin to address subordination at the local level.[117] This suggests an affirmative concept of rights or an anti-subordination principle, enforceable at law, in place of "mainstream equality law," which is "abstract" and "falsely universal."[118]

MacKinnon and West discuss the anti-subordination principle as a means of addressing sexual violence, wife-battering, and prostitution.[119] It is also useful as a way of reframing the issues raised by the forced medical treatment of pregnant women. It eliminates the presumption that a woman's interests can and should be subordinated at some point in time during a pregnancy. And it more accurately describes the conflict as one between the woman and dominant social mores rather than as a conflict between the mother and fetus. The right to choose the course of one's medical treatment is meaningless in the absence of freedom from social coercion.[120] A broad anti-subordination principle aimed at power imbalances rather than at a right to choose except where there are compelling state interests might aid in transforming the standard story to one that includes women.

D. The Limits of a Gender-Only Inquiry

In looking beyond the standard story and to the social construction of gender, we enrich our understanding of patriarchy. This

---

114. MACKINNON, FEMINIST THEORY, *supra* note 6.

115. Robin West, *Jurisprudence and Gender*, 55 U. CHI. L. REV. 1, 70 (1988).

116. MACKINNON, FEMINIST THEORY, *supra* note 6, at 240.

117. *Id.* at 239 ("So long as men dominate women effectively enough in society without the support of positive law, nothing constitutional can be done about it").

118. *Id.* at 242.

119. *See* MACKINNON, FEMINIST THEORY, *supra* note 6, at 247; *see also* Robin West, *Reconstructing Liberty*, 59 TENN. L. REV. 441 (1992).

120. For further discussion of this point see Dorothy Roberts, *Punishing Drug Addicts Who Have Babies: Women of Color, Equality, and the Right of Privacy*, 104 HARV. L. REV. 1419, 1476-81 (1991).

furthered inquiry, however, seems to assume that subordination has conferred a universalizing experience on all women. "Inequality on the basis of sex, women share. It is women's collective condition."[121] The assumption of universality is appealing. It expresses a sense of community, perhaps to galvanize its members to unite in participation in transformative politics. It suggests, however, that the universal experience should define the agenda. It also indicates that elaborating gender on a social level completes the inquiry.

Using "women's collective condition" to define the agenda ignores real and lived distinctions drawn on lines of race, class, and culture. The forms of subordination that will rise to the top will be those experienced or understood by white middle class women. In other words, race, class, and culture will continue to deprioritize issues. The anti-subordination principle will only be implemented against forms of oppression recognized by the dominant class of the subordinated collectivity.

If the inquiry does not progress beyond gender, then we will only learn of and understand gender-based patriarchy. We will only address subordination on a limited basis. The standard story may also dictate this result. Two aspects of the standard story are in evidence here. One is the gender-only inquiry itself. The standard story does speak to gender. It does so by essentializing women, by describing them with regard to the biological capacity to bear children. It uses physical difference to justify social stratification.[122] But none of the reported cases mention race, class, or culture. The furthered inquiry fails in the same way, perhaps blinded by the standard story. The second aspect of the standard story also used by those who tell of socially constructed gender is the notion of universality and collectivity. The standard story assumes only one concept of the social good; it assumes the truth of stereotypes of pregnant women, well-motivated doctors, and of rational, competent decisionmakers; and it assumes that only rights against the state need be defined. And the assumption of universality within the standard story has the same effect as that within the furthered inquiry—more patriarchy.

In theory, a gender-only inquiry could at least eliminate one tier of the hierarchy. But even the most generous interpretation assumes that the subordination experienced by women of color and nonmajority culture can be parceled. That is, the gender-only inquiry seems to consider gender-bias separable from other biases such as race, class, and culture. The experience of women of color does differ

---

121. MACKINNON, FEMINIST THEORY, *supra* note 6, at 241.

122. *See id.* at 242-43 (criticizing the equality debate for focusing on sameness/difference issues). This excludes forms of subordination experienced only by women. For example, "[s]exual abuse has not been seen to raise sex *equality* issues because these events happen specifically and almost exclusively to women as women." *Id.* at 243.

from that of white women, but it is also similar. And it is not separable. A woman of color is not subordinated partly as a woman and partly as a person of color. The intersection between gender and race is more subtle and complex than that. To say that women of color have a particular place in the hierarchy only begins to describe the intersection. Part IV offers a preliminary exploration of intersection.

So perhaps it is more realistic to say that, at the most, the gender-only inquiry might ameliorate the subordination of white, middle class women. Even this appears unlikely. Predicting that a gender-only inquiry could diminish the hierarchy also assumes a fixed number of immutable types of oppression. The experience of domination perpetuates itself. Other forms of oppression will continue to emerge unless we face oppression as a concept and a complex all at once.

IV. CONTINUING THE INQUIRY INTO RACE, CLASS, AND CULTURE

In this section, we must look hard into the silence. The cases at hand do not expressly speak of race, class, or culture. They can address gender directly, even while excluding women's interests and women's voices from the discussion. But the law purports to be blind to color, class, and culture. How is it then that the 1987 survey of obstetricians revealed that of the twenty-one petitions for court-ordered medical treatment, seventeen of the orders were sought against Black, Asian, or Hispanic women?[123] And all of the orders were sought against women being treated at public hospitals or receiving public assistance.[124] The furthered inquiry must ask why pregnant women who are not white and who live in poverty are at a significantly greater risk of forced medical treatment. These women are being measured and found wanting according to non-obvious standards. We must look for those standards and reveal them.

A. Stories About Mothers

The gender-only inquiry does suggest that a standard for good motherhood is being used.[125] The point has already been made that the standard story regulates women as childbearers. The forced medical treatment cases constitute only one part of the regulatory

123. Kolder, *supra* note 9, at 1193.

124. *Id.*

125. *See* Martha L. Fineman, *Images of Mothers in Poverty Discourses*, 1991 DUKE L.J. 274, 289-90 (patriarchy define the concept of "Mother" despite the fact that it is women who live and experience it).

scheme.[126] These cases express at law norms of proper pregnancy behavior. When a court orders the forced medical treatment of a woman for the sake of fetal interests, it deems that woman a bad mother.[127] So, at the least, these cases tell us that a good mother would consent.

We can infer additional details of the model mother by looking to expressions of disappointment within the case reports. The good mother is self-sacrificing and nurturing. This characteristic is revealed by medical staff angered at what they perceive to be the woman's insensitivity to her child's welfare.[128] The courts' willingness to override a decision based on religious beliefs may also express criticism of women who choose religious principle over maternal altruism. The good mother is fearless. Doctors and judges dismiss the concerns of women who are afraid of surgery.[129] And she is compliant. The woman who refuses surgery is regarded as stubborn and irrational.[130]

The good mother is also white and middle class. The silence within the case reports and the loudness of the results indicate that race and class add dimension to the model of motherhood. Within that gap, one can sense negative stereotypes forming a picture of the bad mother.

She has little education. Perhaps she does not understand the nature of her refusal to consent. She is unsophisticated, easily influenced by simple religious dogma. She is pregnant because of promiscuity and irresponsibility. She is hostile to authority even though the state has good intentions. She is unreliable. She is ignorant and foreign. She does not know what is best. The cases ascribe these characteristics to the bad mother; this is the subtext, the things that can nearly be said. They make it easier to assume that the woman's will should be overridden. They also offer moral grounds for intervention. The expressions of anger, frustration, and righteousness in the case reports and opinions strongly evoke the things that can nearly be said. Not stated is that these assumed characteristics are particular to stereotypes of poor women of color. So, what goes

---

126. *See* Dawn Johnsen, *From Driving to Drugs: Governmental Regulation of Pregnant Women's Lives after Webster*, 138 U. PA. L. REV. 179 (1989); Note, *Rethinking (M)otherhood: Feminist Theory and State Regulation of Pregnancy*, 103 HARV. L. REV. 1325 (1990); Note, *Pregnancy Police: The Health Policy and Legal Implications of Punishing Pregnant Women for Harm to Their Fetuses*, 16 REV. L. & SOC. CHANGE 277 (1987-88); Lisa C. Ikemoto, *The Code of Perfect Pregnancy: At the Intersection of The Ideology of Motherhood, the Practice of Defaulting to Science, and the Interventionist Mindset of Law*, 53 OHIO ST. L.J. (forthcoming 1992).

127. Ikemoto, *supra* note 123.

128. *See, e.g.*, Bowes & Selgestad, *supra* note 22 at 211; Leiberman, *supra* note 77, at 515-17.

129. *See, e.g.*, Bowes & Selgestad, *supra* note 21, at 209, 211.

130. *See, e.g.*, Crouse Irving Memorial Hosp., Inc. v. Paddock, 485 N.Y.S.2d 443 (N.Y. Sup. Ct. 1985); Bowes & Selgestad, *supra* note 21, at 211.

unsaid is that she is Black; she is Hispanic; she is Asian; and she is poor.

The act of subordinating occurs first in the mind of those with authority. It is the implicit assumption that women of color, particularly those who live in poverty, are not fit for motherhood. This assumption is rooted in the experience of domination and in telling stories—negative stereotypes—about the "others" to justify the resulting privileged status. Those who call to rights and the feminist critical theorists respond to the good mother model and to the things that can nearly be said. But they overlook the catalyst for subordination in these cases—the mindset of domination with respect to race, class, and culture, the set of assumptions that includes an expectation of conformity, and in the absence of such, an expectation of unfitness.

B. The Culture Clash

A dominant culture of pregnancy and childbirth attaches to the good mother model and effects subordination on race, class, and cultural lines. The forced medical treatment cases illustrate a revealing tension between the dominant culture and the individual women, which reflects a parallel conflict generated by the broader dominant culture.

1. Authoritarianism

The main point of tension is the conflict between the pregnant woman and the institutional authority of medicine wielded by the doctor. The standard legal story acknowledges a conflict between certain individuals and formal government. The story's beginning, the individual right to choose the course of one's medical treatment, describes society as a stronghold of individuals. The invocation of compelling state interests describes a society in which there is significant consensus. This depiction ignores institutionalized authority[131] or authority to which we defer because of a reputational status not based on any one individual and whose sway is greater than the combined weight of the individuals who form the institution. Medicine has become an institutional authority presumed to be a source of valuable knowledge and truth. And it is an institution of privileged knowledge; doctors, the institutional representatives, are presumed to know best. Medicine is also hierarchical. Patients are expected to

131. *See also* Charles Reich, *The Individual Sector*, 100 YALE L.J. 1409, 1411 (1991) (discussing the "Unbalanced Constitition" as the result of the Supreme Court's use of "one approach for organized power and an entirely different approach for the individual").

defer to the greater authority of the doctor. In addition, it reflects the dominant culture in that the privileged few in medicine generally come from the privileged tier of society; it is largely white and male.[132]

Institutional authority has not eliminated autonomy as a goal. Nor is all institutional authority self-privileging. But when it is, it subtracts from our understanding of personal liberty in two ways. First, by its status as a source of knowledge, medicine creates a presumption that there is a better choice to be made—the one that conforms with reasonable medical practice standards. The presumption raises doubt about the authority of the person who chooses otherwise. It changes the decisionmaking process from an opportunity for self-definition to an obligation to meet other-defined expectations. The absence of a presumption would leave more room to understand the person's choice as a moment of self-actualization, rather than as evidence of unfitness. Second, by its disproportional influence, if it does not actually coerce widespread consensus, it gives the appearance of consensus. Thus, the dominant culture, constituted largely of institutions, is largely authoritarian.[133] Deference to authority, institutional or other, is expected. It has become a cultural practice.

We only notice the authoritarian nature of the dominant culture and that of medicine in particular when individuals invoke principles of self-determination and equality to reject this other-authority. Patients who refuse treatment reject the other-authority of medicine. Claimants of self-determination from the inside are more likely to be taken seriously. That is, the patient's claim is legitimate to the extent that he is perceived as aligned with the institutions of authority. The words that can be nearly said here are: Insiders differ less often, so we can assume that they only differ for important reasons; and insiders know and understand the standards, so their rejection of authority is informed.[134] Not given is the description of the inside. The inside is where domination, not subordination, occurs. It is the place where negative stereotypes do not operate. And claims of self-determination will be disregarded to the extent that negative stereotypes have been created to justify domination.[135]

---

132. GENA COREA, THE HIDDEN MALPRACTICE: HOW AMERICAN MEDICINE TREATS WOMEN AS PATIENTS AND PROFESSIONALS (1977); Deloris Kong, *Mds Are White Males—If You Read the Ads*, BOSTON GLOBE, May 24, 1990, at 2.

133. *See* Charles Reich, *The New Property*, 73 YALE L.J. 733, 756-60 (1964).

134. These words that can nearly be said are not conscious excuses, but create and perpetuate an insider reality where privilege is natural. Delgado, *Storytelling*, *supra* note 63, at 2412 ("The stories or narratives told by the ingroup remind it of its identity in relation to outgroups, and provide it with a form of shared reality in which its own superior position is seen as natural").

135. I refer to stereotypes here as both prescriptive and descriptive. That is, they play an active role in the process of subordination, as described above. But the fact that negative stereotypes exist also indicates outsider status.

2. Two Corollaries: Precluding Self-Awareness and Cultural Imperialism

The authoritarian nature of the dominant culture yields two corollaries. One is that privileging institutional knowledge precludes self-knowledge. It disallows women to make self-diagnoses based on their own body-awareness, experience, and beliefs. Most of these cases arise from a woman's refusal to consent to cesarean surgery. Both doctors and women offer a number of reasons for the increase[136] in the percentage of cesarean births.[137] But all of these reasons express a greater concern for physician control than for patient determination.[138] It also indicates that many interventions are medically unnecessary and that the women who felt that labor was progressing normally were right.[139]

The increasing medicalization of pregnancy and childbirth denigrates self-knowledge. In some instances, this denigration is overt. The court in *In re Madyun*[140] granted the hospital's oral petition for cesarean surgery within approximately two and one-half hours. Ay-

---

136. *See supra* note 103 and accompanying text.

137. Marieskind, *supra* note 103, at 197-199 (fear of malpractice suits, heightened use of obstetric technology increased number of women diagnosed as having indications for performing the procedure, economic incentive—"women with the best insurance coverage more frequently have cesareans"); Porreco, *supra* note 103, at 118 (reduce labor time for sake of doctors' busy schedules; perception that birth is an inherently dangerous medical problem, enhances physician's sense of control).

138. "[P]ractice patterns are designed more with the interest of the health professional in mind than the interest of the mothers. . . . [i]t is extraordinarily difficult for people to give up control." Porreco, *supra* note 103, at 122.

139. The 1987 national survey of obstetricians revealed that of fourteen infants delivered by court-ordered cesarean, "[o]nly 2 of 14 infants (14 percent) had important morbidity. . . . No fetal deaths occurred." *See* Kolder, *supra* note 9, at 1193. The authors concluded, that "court-ordered interventions may ultimately cause more problems than they solve. They rest on dubious legal grounds, may expand rather than limit physicians' liability, and could adversely affect maternal and infant health." *Id.* at 1194. *See also* Marieskind, *supra* note 104, at 196 (National studies indicate that approximately one-third of all cesareans are repeat procedures. "Repeat cesareans continue to be performed despite the National Institutes of Health recommendation 'that a proper selection of cases should permit a safe trial of labor and vaginal delivery for women who have had a previous low segment transverse cesarean birth'" and despite the fact that "[t]he safety of the VBAC [vaginal birth after cesarean delivery] in properly selected women has been well documented."); Luis Sanchez-Ramos, et al., *Reducing Cesarean Sections at a Teaching Hospital*, 163 AMER. J. OBSTET. & GYNEC. 1081, 1082 (1990)(reporting results of program to reduce cesarean rate in a largely high-risk, low income obstetric population. "The overall cesarean rate decreased steadily from 27.5% of deliveries in 1986 to 10.5% in 1989 [the study period]." And during the same period, the overall perinatal mortality rate "decreased from 31.8 in 1986 to 14.9 in 1989, . . . whereas the neonatal mortality rate decreased from 16.4 to 6.4." *Id.* at 1084).

140. No. 189-86 (D.C. July 26, 1986). Another court discussed the issue in *In re* A.C., 573 A.2d 1235, 1263-64 (D.C. 1990) (en banc).

esha Madyun had been in labor for about sixty hours, much longer than normal, according to medical standards. The doctor opined that the risk of fetal sepsis increased over time.[141] The Madyuns believed the cesarean was not medically necessary. Nor was Ayesha Madyun permitted to assist delivery herself by walking around. In addition, she asserted "that a Muslim woman has the right to decide whether or not to risk her own health to eliminate a possible risk to her undelivered fetus."[142] But the court issued the order and, in doing so, expressed skepticism at the sincerity of Ayesha Madyun's religious beliefs and questioned the authority of the Madyuns to speak to the issue of choice. "To ignore the undisputed opinion of a skilled and trained physician to indulge the desires of the parents where, as here, there is a substantial risk to the unborn infant, is something the Court cannot do."[143] As expressions of disregard for personal knowledge, these cases may deter many women from seeking such knowledge and trusting their own experience.[144] It may increase reliance on the physician's authority. This limits the likelihood and the person's ability to reject the norm. And in turn, it preserves the institutional authority of medicine.

The other corollary is that privileging norms of exclusive institutions devalues non-mainstream culture. It prefers the strongly risk-averse medical culture of pregnancy and childbirth to the woman-culture developed by the lived experience of women. As discussed, pregnancy is now most often described as a medical event. This can preclude a woman from making a decision based not only on her self-diagnosis but also on facts of life outside the hospital walls. Janet Gallagher reported the case of a Nigerian woman, pregnant with triplets.[145] The medical norm prescribes cesarean delivery for multiple births. The Nigerian woman and her husband refused, in part because the woman was healthy and in part because they planned to return to a region of Nigeria where lack of medical facilities would make future cesareans inaccessible. She understood pregnancy and childbirth in the context of her future. The court, however, focused solely on the medical assessment and issued the order without the woman's knowledge. The medical staff had to tie her down in order to perform the surgery. The effects of this decision include not only three healthy babies but also the physical violation of this

141. *See In re* A.C., 573 A.2d 1235, 1263 (D.C. 1990).

142. *Id.*, at 1260.

143. *Id.* at 1263.

144. SHEILA KITZINGER, WOMEN AS MOTHERS 74 (1978)(describing the standard tests and examinations administered as prenatal care, then concluding, "[o]ne result is that probably the majority of expectant mothers have very little confidence that they are capable of giving birth to a live healthy baby without medical help. They no longer trust their own bodies").

145. Gallagher, *supra* note 41, at 9.

woman and the forcible recharacterization of her understanding of pregnancy—a personal experience implicating health and future—to that of the medical profession—a pathology meriting physician control.[146]

It is also significant that these women were not only outsiders, but they were unfamiliar outsiders. Being Muslim or Nigerian increased the distance between the preferred inside and the devalued outside, and thus increased the justifications constructed for domination and the risks of subordination.

C. Building Choice From Coalition

We gain a more replete understanding of patriarchy by continuing the inquiry. The first lesson was that patriarchy, expressed as the standard story, not only blinds and silences us. It also separates us. If we only respond to the formulae or to gender subordination, we can agree to reclaim rights or we can agree to reformulate the concepts of rights and equality in ways that speak of women's reality. But we cannot agree, without excluding women by race, class, and culture, on priorities and strategies that will affect the lives of each woman. At the most, we can create a thin sense of likeness.

The furthered inquiry does not necessarily reveal a deeper, truer pool of commonality. Even the deeper understanding of patriarchy we gain will not provide a shared experience. Our understandings will differ. But we begin the process of consciousness-raising,[147] and we learn to reject rather than accept justifications for domination. Perhaps we also learn to think of ourselves as a source of promising variety.

We may open doors, through which we may gain insights and form commitments, or whatever is needed to value difference and locate relationships. We may create coalitions and from those, we may build choice.

Opening another door, then, is always the next step. We have looked beyond gender, but we still have not looked to the experience of each subordinated person. The persons most likely to suffer a

---

146. *See* Elizabeth Shearer's commentary in Porreco, et al., *supra* note 101, at 119 ("We do not have sicker mothers or babies; we have a problem in changing cultural and societal attitudes toward birth and the relative roles of women, professionals, and technology in birth . . . Birth is still seen as a medical problem to be managed medically"); R. WERTZ & D. WERTZ, *supra* note 101, at 141. (the increasing role of physicians in childbirth during the twentieth century was accompanied by an increase in intervention *and* a change in the understanding of childbirth as a natural event to an unnatural one. "By 1920 doctors believed that "normal" deliveries . . . were so rare as to be virtually nonexistent").

147. I use consciousness-raising in the way that Professor Matsuda has defined it. "[A] collective practice of searching for self-knowledge through close examination of our own circumstances, in conjunction with organized movements to end existing conditions of domination." Matsuda, *Pragmatism Modified*, *supra* note 5, at 1779.

subtraction of their reproductive rights include not only women of color and culture but also lesbians, gays, and persons institutionalized by criminal conviction or civil commitment. Locating other "have nots" of reproductive choice can help us learn about both patriarchy and possibilities.

CONCLUSION

By continuing the inquiry beyond the standard story and beyond gender, we break the silence and we learn to listen. We learn to understand refusals of consent as moments of dissonance that offer lessons. A woman who refuses consent to cesarean surgery out of fear may have reason to fear. She may know that she cannot provide for the child because she lacks the necessary financial and social support. A woman who refuses consent to a blood transfusion because it violates her religious beliefs may know more than the doctor about her life and her health. Thus, we can see these refusals as moments of integrity, not to glorify, but to respect, and to use to identify real problems— poverty, over-use of medical intervention—not bad mothers.

A FEMINIST ANALYSIS OF PHYSICIAN-ASSISTED DYING AND VOLUNTARY ACTIVE EUTHANASIA

LESLIE BENDER*

Benjamin Franklin said the only certainties in life are death and taxes.[1] As medical technology has developed, our capacity to extend

---

* Professor of Law, Syracuse University College of Law. A version of this paper was originally presented as the Maynard Pirsig Distinguished Lecture, William Mitchell College of Law, in November 1991. I would like to thank the William Mitchell faculty and the law faculties at Syracuse and Cincinnati, the Syracuse Consortium on the Cultural Foundations of Medicine, the Ethics Committee of the Onondaga County Medical Society, the Legal Theory Workshop at the University of Maryland College of Law, and the community at St. Lawrence University for their helpful comments and feedback. My research assistant, Diane Van Epps, Class of 1993, provided invaluable editorial assistance, and Michele Gagnon, Class of 1993, did a wonderful job helping me with the preliminary research.

For my brother, Steven Bender, always caring and attentive to others' needs.

I have titled my essay "A Feminist Analysis of Physician-Assisted Dying and Voluntary Active Euthanasia." I would like to say a little bit about the title. I say "a" feminist analysis because there are many feminist analyses and perspectives, of which my arguments are only one. Feminisms are varied and multiple. Second, I call my work "feminist" because it is grounded in a rich body of writings and thinking in feminist ethics. Because many readers are unaware of the extensive writing and theorizing within feminism and from feminist premises about every kind of subject, they think feminist is a label meaning "political struggles for women's rights." Certainly feminist means that but it also means more.

Some themes in feminist ethics are challenges to the values and conceptions of human natures and human interactions that dominate our current discourses in law, medicine, and ethics. Some feminist theorizing emphasizes the need to value and focus on care, compassion, responsiveness, responsibility, conversation, and communication, as well as learning to listen closely to others and to pay attention to others' needs, regardless of their differences from our own. I write in that tradition.

Feminist ethics also challenges power structures and systemic biases in law and ethics that undervalue or disregard the perspectives and experiences of all women in differing ways and of men of subordinated statuses, whether subordinated by structures of race, class, sexual identity, some other identity-based classification, or some combination thereof. Feminism seeks to reconstruct our understandings and practices in ways that more closely respond to the needs of those people in their daily lives and, I would argue, deaths, or, as I prefer, dying processes. I am as strongly committed to these feminist analyses, although this particular essay, limited by space and time, does not begin to address them adequately.

I have included a bibliography at the end of this essay that collects some of the many writings on euthanasia and feminist ethics that inform ongoing discussions about this topic and bioethics in general—discussions that I hope this essay advances.

1. "But in this world nothing can be said to be certain, except death and taxes." Letter from Benjamin Franklin to Jean Baptiste Le Roy (Nov. 13, 1789), *in* 10 THE WRITINGS OF BENJAMIN FRANKLIN, at 68 (Albert H. Smyth ed., 1905).

human lives beyond what would be their natural deaths has been so astonishing that the keenness of Franklin's aphorism seems nearer to dying than many human bodies. Thus far, we have not mastered "suspended animation" or "immortality,"[2] but medical technicians and scientists have been able to create states of "living" or "undeath" that have not been known before. Lives continue, or are restored, despite hearts stopping, lungs collapsing, livers and kidneys failing, and neocortical brains ceasing to function. For what it is worth, we now can keep bodies "alive" without minds to control them and without any recognizable connection to the personhood or personality of the former owners. Even if science defies the certainty of death, Franklin can still get his due for having it right about taxes.

Regardless of humans' valiant efforts, death remains an unavoidable issue in all our lives. Science and medicine have, at best, learned to delay its inevitability and, at worst, have painfully distorted its processes. Death is seen as a single event, rather than as part of a process of dying. The medical, scientific, and technological segments of our society seem to be in a state of frenzied denial about the inescapable reality of death in everyone's life. Recently, a significant portion of the public has begun to back away from the compulsive drive to extend life at all costs.[3] They have seen its pain, its victims,

---

2. Some people have sought immortality through cryopreservation or cryonics. For example, Thomas Donaldson, along with Alcor, a cryonics organization, unsuccessfully petitioned California courts to permit Donaldson to have his head cryonically suspended before his death and to protect Alcor from prosecution for assisting suicide. *Donaldson v. Van de Kamp*, 4 Cal. Rptr. 2d 59 (Cal. Ct. App. 1992); *see* Miles Corwin, *Tumor Victim Loses Bid to Freeze Head Before Death*, L.A. TIMES, Sept. 15, 1990, at A28; Cynthia Gorney, *Cryonics and Suicide: Avoiding 'the Slippery Slope,'* WASH. POST, May 1, 1990, at D6. The cryonics movement began 30 years ago when Robert Ettinger published THE PROSPECT OF IMMORTALITY (1961). *See* Laura Wisniewski, *Cryonics Groups Pin Their Hopes on the Big Chill*, TORONTO STAR, May 5, 1991, at B6. The title of Ettinger's book indicates the objective of this movement. In 1990, there were already 13 complete bodies and 13 heads in cryonic suspension at three cryonics centers. *See* Maria Goodavage, *Man Pins His Hopes on a Frozen Future; De-Animated—Not Dead*, USA TODAY, Sept. 25, 1990, at 6A. Because it is very expensive to freeze a whole body (about $100,000-120,000), most participants choose to freeze only their heads (at a cost of $28,000-35,000) in liquid nitrogen at 320 degrees Fahrenheit below zero. They hope they can be thawed and cured in the future when there will be the technology to regenerate bodies from the head's remaining cell tissues or to attach other bodies. Corwin, *supra*.

3. *See* Patients Self-Determination Act, Omnibus Budget Reconciliation Act of 1990, § 1866(a)(1), 42 U.S.C.A. §1395cc(a)(1) (West 1992); Cruzan v. Missouri Dep't of Health, 110 S.Ct. 2841 (1990); Bouvia v. Superior Court, 225 Cal. Rptr. 297 (Cal. Ct. App. 1986); Bartling v. Superior Court, 209 Cal. Rptr 220 (Cal. Ct. App. 1984); *In re* Estate of Longeway, 549 N.E.2d 292 (Ill. 1989); *In re* Lawrance, 579 N.E.2d 32 (Ind. 1991); Care and Protection of Beth, 587 N.E.2d 1377 (Mass. 1992); Guardianship of Jane Doe, 583 N.E.2d 1263 (Mass. 1992); Brophy v. New Eng. Sinai Hosp., 497 N.E.2d 626 (Mass. 1986); *In re* Farrell, 529 A.2d 404 (N.J.

its indignity, and its costs. It is not that these people are Luddites who totally reject all medical technology; instead they want technology to be used responsibly and in accord with their needs and values. They are searching for alternative ways to die with dignity, in their homes or with their family and friends, and under circumstances over which they have more control. They are increasingly asking for their physicians to assist them in regaining control over their own dying.

This paper is an attempt to reorient discussions about legal responses to physicians who submit to their patients' pleas for help. I am limiting my discussion here to the "easy" case—to the person who is competent and has expressed her or his wish for physician assistance. Answers to more difficult questions such as those concerning incompetent patients and patients in persistent vegetative states must be reserved for another day. I believe, however, the model I suggest will better enable us to resolve the more difficult questions, but we cannot reach them until we more fully understand the easy case. Even within this simplified inquiry, there are many complex problems that I cannot address here. In choosing which arguments to present, I have focused primarily on offering an alternative feminist legal and ethical paradigm for resolving questions about physician assistance to patients requesting it at the end of life. I realize that switching to a feminist ethic of caring, as I propose, is only part of a feminist analysis. Time constraints have forced me to exclude questions about potential gender biases and gender dynamics within and flowing from application of my proposed model.[4] Although I am omitting these issues from my

---

1987); *In re* the Guardianship of L.W., 482 N.W.2d 60 (Wis. 1992); HASTINGS CENTER OF NEW YORK, GUIDELINES ON THE TERMINATION OF LIFE-SUSTAINING TREATMENT AND THE CARE OF THE DYING (1987); PRESIDENT'S COMMISSION FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, DECIDING TO FOREGO LIFE-SUSTAINING TREATMENT: A REPORT OF THE ETHICAL, MEDICAL, AND LEGAL ISSUES IN TREATMENT DECISIONS (1983); COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS OF THE AMER. MEDICAL ASS'N, GUIDELINES FOR WITHHOLDING OR WITHDRAWING LIFE PROLONGING MEDICAL TREATMENT (1986); George P. Smith, *All's Well That Ends Well: Toward a Policy of Assisted Rational Suicide or Merely Enlightened Self-Determination?*, 22 U.C. DAVIS L. REV. 275, 329 n.392 (1989) (listing several Natural Death Acts); Allan Parachini, *The California Humane and Dignified Death Initiative* HASTING CENTER REP. Jan.-Feb. 1989, at 10-12 (special supp.).

4. I am particularly concerned about three aspects of the gender dynamics and power hierarchies that must be considered in analyzing dilemmas about decisions to end life. First, I am concerned that, thus far, only women have requested assisted death, or at least, the only publicized cases involve women. This may be because women are socialized differently from men and find it easier to ask for help. Additionally, or alternatively, women may find it more unbearable to make significant others suffer from watching their slow, debilitating death. Or women may feel uncomfortable being "cared for" because their socialization usually requires them to be caregivers. Second, there are severe race, gender, and class-based biases in access to health care that must be accounted for in this analysis. Finally, women are primarily the caretakers of the ill and elderly in our society—as nurses, in families, and as health care aides. We also must examine how altering rules about physician assistance in dying affects the experiences, power, roles, and needs of these women.

discussion, I hope it does not minimize their importance, and that you realize this paper is only a small portion of a larger work in progress.

Two more introductory caveats: First, I am deeply committed to major systemic changes in the funding and delivery of health care in our nation. We need a national health care system to ensure no citizen is forced to make medical decisions based on the scarcity of funds or insurance. We also need to re-evaluate the commodification and dehumanization of medicine. The way that most medicine is now delivered, in ten minute segments, often from a series of different physicians or specialists without any sense of continuity for patients, permits few of us to develop relationships with our physicians. It is crucial that we move to a just health care system, although it is impossible without systemic changes in funding and delivery. With that in mind, accompanied by my fears that a revamped health care system is not in our immediate future, my proposal is premised on a continuation of our current system of funding and delivery. That I ask for changes in the system we use should not be understood as an acceptance of this model but as an attempt to work within it until it is changed. Hopefully, my proposals would better serve their ends in a restructured system.

Second, I do not mean to imply that shifting to the paradigm I propose will provide simple answers to all future questions; nor will it eliminate struggles and conflicts in coping with the intersections of dying, medicine, technology, and ethics. It may, however, enable us to deal more humanely, more cooperatively, and more supportively with dying persons, those who love them, and compassionate physicians.

Questions about physician-assisted suicide, active voluntary euthanasia, mercy-killing, or, as I would prefer to call it, "medical care at the end of life" or "medical care in the dying process" have increased over the recent past. Before I make arguments advocating changes in our understandings and legal treatments of this practice, I think it is useful to share with you a brief chronology of some of the most vivid stories shaping this debate in the medical, legal, and bioethical communities.

In March 1991, Dr. Timothy Quill, a Rochester, New York physician, published an impassioned article in the *New England Journal of Medicine*, detailing the decision-making process involved in the death of one of his patients.[5] Diane was a middle-aged business woman, mother, and wife, who was diagnosed as having acute myelomonocytic leukemia, a fatal disease. Dr. Quill advised Diane about available

---

5. Timothy Quill, *Death and Dignity: A Case of Individualized Decision Making*, 324 NEW. ENG. J. MED. 691-94 (1991).

treatment options and the course of her disease without treatment. After careful consideration of her options, Diane refused to undergo painful, drawn-out chemotherapy, which could have given her a twenty-five percent chance of remission. Dr. Quill reported being very troubled by his patient's decision, but he acknowledged the decision about treatment was hers to make. He had no doubt that she made it in an informed, rational manner with her husband and son.

Once she had decided to forego potential life-preserving treatment, she had to decide how to cope with the disease, her pain, and her inevitable, rapidly approaching death. Diane was advised about hospice and comfort care treatment. She asked Dr. Quill about help in dying. Because Dr. Quill was conflicted about determining his appropriate role in responding to this request, he recommended that Diane contact the local Hemlock Society to learn more. She did and later returned to Dr. Quill complaining of sleeping troubles and asking for a prescription. Knowing that this was part of a Hemlock Society "suicide," Dr. Quill questioned her about how she would use the drugs. He described the interaction as follows:

> In our discussion, it was apparent that she was having trouble sleeping, but it was also evident that the security of having enough barbiturates available to commit suicide when and if the time came would leave her secure enough to live fully and concentrate on the present. It was clear that she was not despondent and that in fact she was making deep, personal connections with her family and close friends. I made sure that she knew how to use the barbiturates for sleep, and also that she knew the amount needed to commit suicide. . . . [S]he promised to meet with me before taking her life, to ensure that all other avenues had been exhausted. I wrote the prescription with an uneasy feeling about the boundaries I was exploring—spiritual, legal, professional, and personal. Yet I also felt strongly that I was setting her free to get the most out of the time she had left, and to maintain her dignity and control on her own terms until her death.[6]

Three and one-half months later, Diane's condition had deteriorated, and she was left to choose between increasing pain and discomfort or sedation and dependence. Diane then called Dr. Quill and her friends to say goodbye. When she decided her life was over, she asked her husband and son to leave her alone for an hour, and she died peacefully on her couch at home. Dr. Quill called the medical examiner and reported her cause of death as "acute leukemia." By sharing Diane's story and his own angst in caring for her, Dr. Quill made a

---

6. *Id.* at 693.

heart-felt plea to the medical community for more dialogue about the appropriate roles of physicians in helping their patients achieve death with dignity.

Dr. Quill's plea to the medical community was quickly transferred to the legal community when a Rochester area District Attorney, Howard Relin, said he would attempt to prosecute Dr. Quill for assisting a suicide—a crime in New York. Initially, Relin's inability to identify "Diane" impeded prosecution, but an anonymous phone call gave him enough information to proceed.[7] The case against Dr. Quill was presented to a grand jury in July 1991, which, in its wisdom, refused to indict him.[8] Dr. Quill suffered the threat of criminal prosecution for over four months—not a pleasant experience for a doctor who conscientiously and compassionately cared for a dying patient.

Dr. Quill's story is just one of a number of stories about physician-assisted death and the legal system's response. Several physicians have been prosecuted over the years, and others have avoided prosecution by keeping their actions secret from the public and sometimes even from the families of the patients they helped die. Other physicians, like Dr. Quill, have sought publicity to raise the public's consciousness on the issue. In some ways, Dr. Quill's story was an antidote to powerful, contemporaneous media stories about Dr. Jack Kevorkian, a retired Michigan pathologist, who invented a "suicide machine" to enable dying patients to voluntarily end their lives once the suffering or loss of dignity accompanying their diseases became unbearable. In 1990, Dr. Kevorkian's actions captured the public and medical ethicists' attention when he permitted Janet Adkins, a fifty-four year old woman suffering from Alzheimer's disease, to use his machine to end her life.[9] The apparent publicity-seeking nature of Dr. Kevorkian and the strange conditions of Adkin's death—that is, the use of this jury-rigged machine in the back of a 1968 Volkswagen van—had skewed the debate about physician-assisted death. Dr. Kevorkian ultimately avoided criminal prosecution in December 1990 because Michigan did not have a law criminalizing the assistance of suicide, and his actions did not amount to murder since he did not cause her death.[10] He was subjected to a civil court order in early 1991 enjoining his further use

---

7. B.D. Colen, *On Death and Dying—MD Who Aided in Suicide Aims to Humanize Debate*, NEWSDAY, Aug. 11, 1991, at 3.

8. *Id.*

9. Kevorkian named his crude, jury-rigged contraption Thanatron, but his lawyer defense team suggested he rename it the "Mercy Machine" for trial. Ron Rosenbaum, *Angel of Death: The Trial of the Suicide Doctor*, VANITY FAIR May 1991, at 147. Kevorkian now refers to his machines as "mercytron." What is in a name? *See infra* note 21-27 and accompanying text.

10. *Case Against 'Dr. Death' Dropped After MI Judge Throws Out Charge*, 7 MED. ETHICS ADVISOR 13-16 (1991).

of his machine or one like it to aid someone else in committing suicide.[11] Last October, while he was still in the process of appealing that order, Dr. Kevorkian assisted two other women in ending their lives.[12] He is now being threatened with contempt proceedings for violating the injunction. His medical license was revoked on November 20, 1991, and on February 5, 1992, he was arrested after a grand jury returned two murder indictments and criminal charges for illegal delivery of a controlled substance against him.[13]

Nonetheless, in March 1991, when Dr. Quill's article was published, his presentation of the issues recaptured the public debate and returned it to a seemingly more medicalized model. Because I am a teacher of bioethics and law, these stories instantly seized my attention. They have also affected political and legislative processes. Washington citizens, through the initiative process, introduced a Death with Dignity proposal on their November 1991 ballot, which, had it passed, would have been the first in the country to explicitly permit physician aid-in-dying.[14] The measure failed by a small margin at the polls (fifty-four percent to forty-six percent), despite earlier polls that indicated a sure victory.[15] California citizens are currently attempting to get a slightly different referendum regarding physician-assisted death on the 1992 ballot—one providing for a waiting period and family notification.[16] Citizens in Oregon and Florida are pressing initiatives for 1994 and 1996, respectively.[17] Other states are also engaged in these debates.[18] Additionally, Derek Humphry, founder of the Hemlock So-

---

11. Judge Alice Gilbert of the Oakland County Circuit Court in Michigan, before whom one facet of Kevorkian's case was heard, described Dr. Kevorkian as having "a propensity for media exposure and seek[ing] recognition through bizarre behavior." *Kevorkian Told: Hands Off Machine!* 4 DOCTOR'S PEOPLE NEWSLETTER 2(1) (March 1991). His lawyers have appealed her decision to enjoin use of his machines and chastised the judge for her moralizing and "unprofessional attack" on Dr. Kevorkian. *Permanent Ban Against Assisted Suicide Appealed*, UPI, Feb. 22, 1991, *available in* LEXIS, Nexis Library, UPI File.

12. Isabel Wilkerson, *Opponents Weigh Action Against Doctor Who Aided Suicides*, N.Y. TIMES, Oct. 25, 1991, at A10; Eric Harrison, *"Dr. Death" Arrested in 2 Women's Suicides*, L.A. TIMES, Feb. 6, 1992, at A15.

13. *Kevorkian Chronology*, Gannett News Service, Feb. 5, 1992, *available in* LEXIS, Nexis Library, Gannett File. In May 1992, Kevorkian assisted Susan Williams's death by giving her canned carbon monoxide. Al Koski, *'Dr. Death' Strikes Again*, UPI, May 16, 1992, *available in* LEXIS, Nexis Library, UPI File.

14. Washington Citizens for Death with Dignity led the movement for passage of Initiative 119, which provided for "aid in dying" as a right of terminally-ill, mentally competent patients. Joyce Price, *Ire Over Prosecution Helps 'Right-to-Die' Bill*, WASH. TIMES, May 13, 1991, at A4; Merle S. Goldberg, *The Right to be Right; Ethics Issues Grow in Number and Complexity*, WASH. TIMES, June 3, 1991, at M3.

15. Jane Gross, *The 1991 Election: Euthanasia; Voters Turn Down Mercy Killing Idea*, N.Y. TIMES, Nov. 7, 1991, at B16.

16. Janny Scott, *Suicide Aid Focus Turns to California*, L.A. TIMES, Nov. 7, 1991, at A3.

17. *Id.*; *see also supra* note 14.

18. Proposed measures in the United States are often compared to those in

ciety, a group organized around the right to die, recently published *Final Exit*, a book of explicit instructions on suicide methods directed toward the terminally ill.[19] The book has been a best seller since its release. Sales have been impeded only by a shortage of printed copies.

This brief summary of some recent stories about death and dying in our culture raises many questions for lawyers. How ought our legal system respond to these human experiences of dying, medicine, and technology? What role should law play in resolving these foundational ethical dilemmas? More particularly, as a lawyer and academician interested in bioethics, I want to examine how law ought to deal with physicians who assist their patients in ending their lives. Soon after Dr. Quill's story reached the press, I began my research on laws about suicide, assisting suicide, and homicide. I learned that no longer do any states have laws criminalizing suicide, and about half the states have laws making assisting or causing suicide a crime, but my questions about what the law ought to do and why remained unanswered. I then read ethicists, philosophers, physicians, and legal scholars' writings on euthanasia, suicide, and terminal illnesses. I can assure you there is more material out there about this area of inquiry than any one person could ever read in a lifetime.[20]

---

the Netherlands, where doctor assisted death, or voluntary active euthanasia, is excused, if strict guidelines are followed. In the Netherlands, the courts and medical societies have established guidelines for when active, voluntary euthanasia by physicians will be legally justified (that is, not subject to criminal prosecution). *See generally*, John Horgan, *Science and the Citizen: Death with Dignity*, 264 SCI. AM. 17 (1991). The Royal Dutch Pharmacists Association has published a ten-page pamphlet explaining the most sensible ways for physicians to offer their patients "death on request." Michael Specter, *Thousands of Dutch Choose Euthanasia's Gentle Ending; U.S. Physicians Debate Death on Request*, WASH. POST, Apr. 5, 1990, at A1. Doctors' death-assisting conduct is evaluated after the fact by local prosecutors who determine if the guidelines, which refer to things like repeated patient requests, medical consultations, and interminable suffering, were complied with, in which case no prosecution follows or a finding of not guilty will be entered. *See id.* When doctors fail to follow guidelines but death assistance is compassion motivated, a guilty verdict without punishment may result. *See id.* For a carefully detailed examination of the development of active voluntary euthanasia in the Netherlands, see Wainey, *infra* note 21 at 653-64. Many commentators have suggested that the United States follow a model similar to that in the Netherlands. *See, e.g.*, George Garbesi, *infra* note 21. It would be unfair, however, to represent the Dutch system as without serious dissent. Because of the tensions involved, Ineke Stinissens, a 47 year old woman who had been comatose for 15 years because of an overdose of anesthesia during childbirth, was forced to starve to death for 11 days after her feeding tubes were removed. Galina Vromen, *Patient's Starvation Death Intensifies Dutch Mercy-Killing Row*, The Reuter News Reports, Jan. 20, 1990, *available in* LEXIS, Nexis Library, Reuter File. Her husband spent years in court trying to get her nursing home to let her die, but when a court finally agreed to let her tubes be disconnected, it refused to order doctors at her convalescent home to end her life. *Id.* Issues of euthanasia are affecting political coalitions in the Netherlands.

19. DEREK HUMPHRY, FINAL EXIT: SELF-DELIVERANCE AND ASSISTED SUICIDE FOR THE DYING (1991).

20. For a comprehensive compilation of resources, see Smith, *supra* note 3;

The more I read and thought, the more it seemed that my questions about the role of law in these situations could not be answered without a final determination about the meaning of life, the meaning of personhood, and the meaning of death. As I drifted in a sea of philosophical and spiritual inquiry, Dr. Quill and Diane's story kept calling me back. This was not a hypothetical problem or an abstract investigation. The legal system's treatment of this problem was affecting the resolution of these dilemmas in people's lives every day, whether I ever figured out the propriety of using a sanctity-of-life or quality-of-life analysis, whether I could discern under what circumstances suicide was ever morally justified, or whether a meaningful difference between "letting die" and "killing" existed. Questions about the legality of physician-assisted death concerned real people, immediate dilemmas, and intense suffering.

My critical feminist consciousness was aroused. Why were the debates framed in terms of abstract principles like autonomy, paternalism, and beneficence or revealed through abstract, hypothetical situations perched tenuously on slippery slopes—for instance, some theorists argue that if we let a doctor respond to a request by a terminally ill patient to die, this would lead to doctors killing disabled people or the elderly poor against their will. Why did the ethical or legal analyses seem to emphasize labeling actions as suicide, assisted suicide, euthanasia, murder, or refusing treatment rather than to emphasize examining the specific facts and contexts, discussing people's feelings and relationships, and responding to patients' needs? Why do we discuss informed consent and who should decide, without discussing caregiving, compassion, responding to needs, interpersonal relationships, dignity, empowerment, and love? And last, but not least, why are we always very careful to leave out the needs and interests of family, friends, and caregivers when we discuss a dying person, as if those subjects are taboo?

Feminist theories help me in my inquiries because they press me to question assumptions and labels and to eschew universal rules, abstractions, and generalizations that impede attention to contexts and lived experiences. Feminist theories promote the values of caring, responsibility, and responsiveness to needs absent in our current legal paradigm.

I. Power Of Naming

An elementary premise of feminist theories recognizes that defining or naming a problem is a political act.[21] When we call doctors' actions

---

Don V. Bailey, The Challenge of Euthanasia: An Annotated Bibliography on Euthanasia and Related Subjects (1990).

21. One of the most difficult parts of discussing this issue is naming the

"aiding suicide," "euthanasia," or "killing," we prefigure the ensuing debate. These labels carry pejorative baggage. Suicide is often connected with notions of irrationality and wrongdoing, whether a spiritual or religious wrong or a mistake in judgment that ought to be corrected. When we think someone behaves nobly in consciously sacrificing her life, we do not label her act suicide. If someone throws himself before an oncoming car to prevent his child from being hit, we do not say he committed suicide. If a fire fighter dies putting out a fire, we do not label her act "suicide."

Freud committed suicide by asking his physician to end his suffering from cancer of the jaw.[22] Bruno Bettelheim committed suicide.[23]

---

subject area. The names used for this phenomenon are wide-ranging, and each label shapes the discussion in a particular way. Institute of Medical Ethics Working Party on the Ethics of Prolong Life and Assisting Death, *Viewpoint: Assisted Death*, 336 LANCET 610, 611 (Sept. 8, 1990) [hereinafter *Viewpoint: Assisted Death*]; *see also* MARTHA MINOW, MAKING ALL THE DIFFERENCE (1990) (discussing the power of naming and labeling and its effects on our thinking). Theorists of physician-assisted death have recognized the relevance of naming by using varieties of labels: Physician- or doctor-assisted death, *Viewpoint: Assisted Death*, *supra*; rational suicide, Stephen A. Newman, *Euthanasia: Orchestrating "The Last Syllable of . . . Time,"* 53 U. PITT. L. REV. 153, 161 (1991); euthanatic rational suicide, Shari O'Brien, *Facilitating Euthanatic, Rational Suicide: Help Me Go Gentle Into That Good Night*, 31 ST. LOUIS U. L.J. 655 (1987); assisted suicide, George C. Garbesi, *The Law of Assisted Suicide*, 3 ISSUES L. & MED. 93, 93-111 (1987), H. Tristram Engelhardt, Jr. & Michele Malloy, *Suicide and Assisting Suicide: A Critique of Legal Sanctions*, 36 SW. L.J. 1003 (1982), Victor G. Rosenblum & Clarke D. Forsythe, *The Right to Assisted Suicide: Protection of Autonomy or an Open Door to Social Killing?*, 6 ISSUES L. & MED. 3 (1990); mercy-killing, James S. Goodwin, *Mercy Killing: Mercy for Whom?*, 265 JAMA 326 (1991); perimortal initiatives, Count D. Gibson, *Perimortal Initiatives: Issues in Foregoing Life-Sustaining Treatment, Suicide, and Assisted Suicide*, 3 ISSUES L. & MED. 29 (1987); timing-of-death decisions, Sandra Segal Ikuta, *Dying at the Right Time: A Critical Legal Theory Approach to Timing-of-Death Issues*, 5 ISSUES L. & MED. 3, 3-66 (1989); life-shortening palliative care, Donald G. Casswell, *Rejecting Criminal Liability for Life-Shortening Palliative Care*, 6 J. CONTEMP. HEALTH L. & POL'Y 127 (1990); aid-in-dying, *Model Aid-in-Dying Act*, 75 IOWA L. REV. 125 (1989); enlightened self-determination, George P. Smith, II, *All's Well That Ends Well: Toward a Policy of Assisted Rational Suicide or Merely Enlightened Self-Determination?*, 22 U.C. DAVIS L. REV. 275 (1989); arranged or negotiated deaths, Catherine Shaffer, Note, *Criminal Liability for Assisting Suicide*, 86 COLUM. L. REV. 348, 369, 370 (1986); consciousness, Steven Goldberg, *The Changing Face of Death: Computers, Consciousness, and Nancy Cruzan*, 43 STAN. L. REV. 659 (1991); direct and indirect euthanasia, JOSEPH FLETCHER, HUMANHOOD: ESSAYS IN BIOMEDICAL ETHICS 149 (1979), Robert Barry & James Maher, *Indirectly Intended Life-Shortening Analgesia: Clarifying the Principles*, 6 ISSUES L. & MED. 117 (1990); active voluntary euthanasia, Helga Kuhse, *The Case for Active Voluntary Euthanasia*, 14 LAW, MED. AND HEALTH CARE 145 (1986), Deborah A. Wainey, Note, *Active Voluntary Euthanasia: The Ultimate Act of Care for the Dying*, 37 CLEV. ST. L. REV. 645 (1989); active euthanasia, Francis Molenda, *Active Euthanasia: Can It Be Justified?*, 24 TULSA L. REV. 165 (1988).

22. In 1939, Freud asked his doctor to inject him with sufficient drugs to kill him when he could no longer bear the suffering from incurable cancer of the jaw.

So did the eminent jurist, Judge Henry Friendly.[24] So, arguably, did Socrates.[25] Yet we rarely talk about rational suicide because the words seem incongruous. Many of our legal opinions are carefully crafted to distinguish between suicide and decisions to forego medical treatment. When a patient asks that a ventilator be withdrawn or refuses consent for chemotherapy, the law does not say that the patient is committing suicide. Jurists have recognized the need to distance these acts from that label. An act or decision is not inherently suicide. A social context is needed to understand the act, making suicide a social construct. At least for the time being, the social construction of that word imbues the event or act with a taint of illegitimacy. State laws making it criminal to aid a suicide are based on underlying assumptions that suicide is irrational or the aid represents a form of coercion. Therefore, if we label the doctor's conduct as "aiding suicide," we raise up those senses of wrongfulness, irrationality, or coercion, when they may, in fact, be totally absent from the event.

Likewise, the word "euthanasia," while originally meaning "good death,"[26] has also taken on images of coercion. Despite careful distinctions in law and ethics between voluntary and involuntary euthanasia, or between murder and euthanasia, mention of the word raises the specter of Nazi Germany in too many minds. The word is infused with imagery of forced exterminations and immoral medical practices. It almost seems that many contemporary physicians and bioethicists think euthanasia is a German word meaning gas chambers, lethal injections, and selective exterminations. Hence, discussing this problem as one of active voluntary euthanasia brings forth debates about slippery slopes and bad actors doing immoral acts.

The words "killing" or "kill" color the discussion even more negatively. Killing sounds criminal. To say it is "mercy killing" for a physician to give a patient a requested lethal dose to end her suffering is to invest the act with an air of criminality that the word "mercy" does not adequately temper. Words affect how we think and feel about acts, how we classify them, and how we treat them legally.

Understanding the politics of naming, as any feminist lawyer does, I prefer to discuss this "issue" as one about death, not suicide,

---

Victor Cohn, *An Assisted Suicide; Is it the First Step Toward Euthanasia?*, WASH. POST, June 12, 1990 (Health), at 27.

23. Celest Fremon, *Love & Death; In His Final Interview, Just Before His Suicide, Bruno Bettelheim Explained Why He Wanted to Die*, L.A. TIMES, Jan. 27, 1991 (Magazine), at 17.

24. *Henry Friendly; Judge Was "One of the Greatest,"* CHI. TRIB., Mar. 13, 1986, at C8.

25. *See, e.g.*, ROBERT CAMPBELL AND DIANE COLLINSON, ENDING LIVES 8-12 (1988).

26. The prefix "eu" means well or good and "thanatos" means death. WEBSTER'S NEW UNIVERSAL UNABRIDGED DICTIONARY 631 (2d ed. 1983).

euthanasia, or mercy-killing. Yet even death is inadequate because it seems to indicate one moment or event rather than an ongoing process. A better naming would be to say this is about "dignified dying," or about "timing of dying decisions," "end of life decisions," "care for dying people," "controlling our own dying process," or "life-completing decisions." Certainly we do not object to giving people control over the completion of their lives. Few, if any, would be morally offended by the provision of care for dying people.

There are other phrasings of these problems that subvert constructive conversations. I am troubled by use of the phrase "physician-assisted" or "doctor-assisted." By starting with the physicians or doctors, we center our attention on them, even though they ought not be the focus. By labeling the event "physician-assisted dying," we concentrate on the actions of the physician, almost making it sound as though the physician is a decision-maker, rather than orienting ourselves toward a discussion of the entire decision-making process and the dying person. And while I am seeking new labels, I would like to reject the word "patient." As the hospice movement has so aptly discerned, the word "patient" has come to connote passivity—someone acted upon.[27] It is objectifying and distancing. Certainly we can find a better word with more decision-making agency and more subjectivity.

II. False Dualisms

Feminist and post-structuralist theories have also criticized our tradition of viewing the world in dichotomies—seeing events as polar opposites; drawing lines that divide the world of concepts into twos.[28] It is not just our naming of things that is problematic but also our narrow bipolar classification schemes. I would like to highlight a few examples of how this has impeded our ability to work through problems of physician assistance to dying people. Dualistic thinking leads us to an either/or, self/other analysis instead of plural, multiple, variant, and contextualized analyses. While dividing all things into two groups, where some thing or event must fall in or out of the group, simplifies our tasks of classification, it papers over the ever-changing relationships and interconnections between categories and experiences and deludes us into believing the categories are fixed, natural, or inherent.

As technology has advanced, we have learned that our understanding of life and death as opposites and fixed categories is inadequate.

---

27. Hospice calls the people with whom they work clients instead of patients. Alice Lind, *Hospitals and Hospices: Feminist Decisions about Care for the Dying, in* Healing Technology 263, 270 (Kathryn S. Ratcliff ed., 1989).

28. *See, e.g.*, Frances Olsen, *The Sex of Law, in* The Politics of Law 453 (David Kairys ed., rev. ed., 1990).

Life and death seem more like interrelated processes on a continuum than clearly delineated and oppositional states. Although our culture holds fast to a view of science as distinct from faith, we continuously encounter difficulties differentiating science from spirituality, especially when we analyze birth and death issues. Nonetheless, our dominant cultural norms seem to privilege doctors' medical knowledge over others' knowledges in caring for dying people, as if death and dying were purely "scientific" or "medical" processes. Despite the ill-fitting nature of our bipolar categories, we continue to separate and favor science over spirit, body over soul, reason over emotion, and self over others. An ontology that examines relationships and interactions among concepts, actions, people, and institutions seems preferable to one fixated on delineating boundaries.

Another false dualism that dominates discussions about euthanasia in law and ethics is active/passive or act/omission. Whether we call something "active" or "passive," or "killing" or "letting die," it is a conclusion, not an inherent fact. Criminal laws and tort laws often distinguish between acts and omissions. If a doctor gives a patient a lethal injection, it is considered an affirmative act, resulting in a charge of active euthanasia or homicide. Yet, if a doctor withdraws a life-support system, whether a respirator or a feeding tube, that is "letting nature take its course," an "omission," or being "passive." Even though there is an "act" of detaching, it is not considered active. Even though the need to detach is related to the earlier "act" of attaching the person to life-prolonging machinery, it is considered merely passive or omissive to withdraw it. An original decision not to attach a patient to life-support, whether due to triage resource allocations or apparent futility, is also passive. There is no legal liability and, for most ethicists, no ethical liability. In each case, the patient dies in conjunction with a decision about her care that is effected by a physician. Yet when a person dies in conjunction with a decision to end her suffering from a terminal illness that is effected by a physician giving her a lethal injection or a prescription for a potentially deadly dose of medicine, the law seems to say it falls on the killing side of the dichotomy.

Once an appropriate decision to complete the life process and allow death to occur has been made, physicians, ethicists, and the legal system should seek out the most compassionate way to care for the dying person. It is unseemly for the legal system's analysis to turn on whether the physician's role was active or passive, or whether the conduct is more appropriately labeled killing or letting die. Many prominent theorists have argued against this distinction's relevance much better than I can.[29] They challenge legal and ethical paradigms

29. *See, e.g.*, James Rachels, *Active and Passive Euthanasia* in EUTHANASIA: THE MORAL ISSUES 45-51 (Robert M. Baird & Stuart E. Rosenbaum eds., 1989).

that evaluate physicians' conduct through an active/passive, act/omission lens.

Physicians are not immune from this dualistic approach to assessing active voluntary euthanasia. Many physicians measure appropriate conduct through a healing/killing dichotomy. Doctors say they are trained to heal, not kill, as if those terms covered the whole universe of actions, as if they are fittingly contrasted, and as if actions could clearly fall in one category or the other. I have heard doctors claim that the bright line between healing and killing is necessary to keep physicians principled and honorable. If the law permits a blurring of the line, they argue, doctors may prematurely end the lives of dying or obstreperous patients out of impatience or exhaustion from the heavy demands of caring for a desperately ill, dying patient, out of frustration or a sense of defeat at their lack of success in curing the patient, or even for economic reasons.

The law seems to use similar justifications for its active/passive or killing/letting die distinctions. These rationales are legitimate only if we agree with three underlying assumptions: 1) laws and ethical principles must be designed for the "bad actors"; 2) each line must be firmly set to prevent a precipitous decline down the proverbial slippery slope; and 3) truly bad actors are in fact deterred by laws. I am unpersuaded by each. Although there are, and always will be, a number of bad actors, most of us do not fall in that category. If we write our laws or set our standards to curtail the actions and improper motivations of a small contingent of people on the margin, we may disempower the majority of us in the center from acting on noble and virtuous impulses. Physician aid-in-dying exemplifies this critique.

To deter negligent, indifferent, malevolent, or lazy physicians from involuntarily terminating some patients' lives (or wrongfully persuading patients that death is their only option), we have to endure a rule that deters compassionate physicians from providing competent, suffering patients requested dignity, security, and control over their dying processes. Similarly, we prohibit family members from mercifully ending the suffering of loved ones or create high legal barriers to families making termination of life-support decisions for incompetent loved ones based on our fear of bad families. The social and ethical price of designing our laws and rules for the bad actors is significant suffering and indignity to innocent, humane people because of unnecessary restraints on their freedom to act out of care in a manner responsive to particularized circumstances of need.

Laws making doctor assistance illegal may deter caring physicians from acting. Few doctors want to be vulnerable to the whims of prosecutorial discretion, particularly if it is an election year, and even fewer want to risk the possibility of criminal prosecution or license revocation, although they are likely to prevail ultimately. Compassionate and caring doctors who want to comply with their patients' pleas will be deterred, unless we have laws clearly authorizing them to act

and outlining conditions under which they will be free from prosecution.

Why are we so quick to constrain the power of most people's moral agency? Why do we presume that if we give physicians freedom to implement their patients' decisions about care at the end of dying, they will behave irresponsibly? As a society, we readily give physicians a great deal of responsibility to exercise their best judgments and skills in caring for patients. If we are willing to presume they are responsible enough under most situations to deal with matters of life and death, why would they suddenly be less responsible in helping to implement patients' decisions at life's end?

Moreover, I am not completely convinced that such laws or rules are very effective deterrents to the truly bad actors—the lazy or callous physicians or parsimonious families who are cruel to those in their care. Despite the existence of these rules, we still have bad actors who violate them. Clearly some in the small group of bad actors in the margin are never deterred by laws. For this reason, the class of people for whom we are calibrating our laws is reduced in size even further. Maybe we should reconsider whether the cost of pitching our laws to this relatively small number of people, at least in cases of aid-in-dying, is too great for the benefit we receive. If we respect the autonomy and dignity of dying people, we should make laws that create an environment where people can get the care they need from their physicians, rather than laws that merely deter a few dishonorable, bad physicians.

Our legal system loses its legitimacy when faced with questions of doctor-assisted death. Laws seem to make doctor-assisted death criminal, based on active/passive distinctions or notions of irrational suicide, and yet doctors are rarely prosecuted and even more rarely convicted.[30] The law says it is impermissible, but then winks at the conduct. Prosecutors often use their discretion not to prosecute, and juries use their discretion to dismiss acts of mercy.[31]

In one way, one could say that our legal system is responding appropriately. It is contextualizing our system of justice to fit the circumstances. To that extent, these verdicts or results excusing physician conduct are good. But there are other dynamics about which we need to be concerned. If the active/passive distinction is a correct ethical and legal analysis, then juries and prosecutors ought not subvert the law. If there is something fallacious or ill-fitting about the active/

30. For detailed reviews of earlier cases against physicians, see Wainey, *supra* note 21, at 668-70 (Drs. Sander, Montemarano, Kraai, Neidjil, Barber, Hassman, Rosier and Caraccio); DEREK HUMPHREY & ANN WICKETT, THE RIGHT TO DIE: UNDERSTANDING EUTHANASIA (1986) 42-45 (Dr. Sander), 103-04 (Dr. Montemarano), 140-42 (Dr. Kraai) (1986); *Eight Doctors On Euthanasia Charges*, HEMLOCK Q., Jan. 1989, at 6.

31. *Id.*

passive distinction, then we should find a better analysis for judging the legality and rightness of compassionate acts complying with patient requests to end their lives. More often than not, juries reject our present model because it does not reflect their experiences and understandings of justice. If, for the most part, our legal system is clandestinely applying an ethic of care in these cases, why not bring it out in the open? It would not be a radical shift because it represents the current practice, if not the language, of the legal system. To permit the laws to be overtly disrespected by judges, prosecutors, and juries impairs the legitimacy of our legal system.

If what we are talking about is physician participation in the care of dying people, it should not matter whether a physician helps by disconnecting machines, by giving an injection, or by giving a prescription. The appropriateness of the conduct should not turn on an artificial distinction between healing and killing. What should matter, and what we should be asking about, is whether a physician thoroughly discussed the medical aspects of the dying process and care options with the dying person, and whether there have been ongoing conversations about dying between the dying person and loved ones, caregivers, and medical providers. We then should ask whether the physician was "giving medical care" that responded to the dying person's needs, concerns, and values.

III. A Caring Paradigm For Medical Ethics And Law

It is this notion of "giving needed medical care," informed by an ethic of care paradigm, that I want to explore for the rest of this paper. My feminist critique of medical ethics and legal practice regarding this issue is ultimately a critique of the paradigm we use. In part because of the language usage and dualism problems I discussed, but more because of our dominant, liberal paradigm premised on a society composed of autonomous individuals who interact with others by choice out of self-interest, we look for resolutions of problems about end-of-life medical care in an ethic of justice and rights. We construct abstract, generalized rules that are supposed to cover all situations for all time. Our current analysis prevents people from aiding others to die with dignity because we understand rights as barriers to interference by others, rather than as enabling conditions. Our ethical constructs grow out of elaborate conversations, which are deeply philosophical and richly argued, and yet we leave out the heart and soul of real people's concerns about dying. We leave out discussions about caring, empathy, love, compassion, relationships, and the dying person's needs and perspective. When applying our existing rules to the legality of physician assistance in the dying process, we may talk of "mercy seasoning justice," but I would prefer an understanding that speaks of "justice tempering care." We can change the substance of our normative discourse in medical ethics and law by moving to a care-based paradigm like the one I propose.

In addition to the absence of values like compassion and care, and the focus on rights as barriers between independent equal individuals, our current ethical paradigm is defective because it fails to account for the effects of changes in technology on analyzing issues of dying. Medical ethics, for instance, is ahistorical because it relies on ancient ethical codes, such as the Hippocratic Oath,[32] devised 2500 years ago in a a wholly different historical and social context in which our present medical technologies were unfathomable. Many medical ethicists and doctors follow these codes and declare that active killing of dying patients is wrong, regardless of the circumstances.

Modern society is characterized by a boundless quest for technological innovation to dominate nature and control life processes. Our drive for technological mastery of natural phenomena has often eclipsed our humanitarian and ecological concerns. In medicine, technology has been as wonderful as it has been alienating and destructive. Sometimes our strivings for medical and technological glory and for conquering death are so strong that we lose sight of the suffering we prolong and create.

Even where technology has succeeded in fending off death's assaults, it often distances us from the feelings and experiences of those who are dying. People seem less touchable, less human, and less real when connected to complicated medical equipment and tubing. They are often in intensive care or special hospital units, blocked off from visitors and all things familiar. Our technological revolution in medicine has usurped many people's opportunities to die with dignity at times or in manners of their own choosing, with their family or friends around, and in their homes.

Concepts of justice and rights should not be jettisoned when shifting to an alternative feminist analysis, but they should be used as correctives to an ethic of care when needed to make sure that power is not abused. A care- and responsibility-based ethic rests on assumptions that seem closer to the experiences of dying and death in people's lives than assumptions underlying a rights- or rule-based ethic, which arguably might be more appropriate in other settings. A care-based ethic arises out of perceptions of human beings as relational, interdependent, and supportive as opposed to our current rights-based ethic in which people are separate, autonomous, and equally empowered actors. A care-based ethic acknowledges that emotions are as important as reason in our lives, decision-making, and dying, and that preserving relationships with and enabling others is as important as having rights to protect us from others.[33]

---

32. L. Edelstein, *The Hippocratic Oath: Text, Translation and Interpretation*, 19 BULL. HIST. MED. 1164 (1943); *Hippocratic Oath*, in JUDITH AREEN, ET AL., LAW, SCIENCE AND MEDICINE 273 (1984); *see* Curley Bonds, *The Hippocratic Oath: A Basis for Modern Ethical Standards*, 264 JAMA 2311 (1990) (arguing that the ancient oath's fundamentals are still applicable today).

33. *See generally* CAROL GILLIGAN, IN A DIFFERENT VOICE (1982). Some

Feminist ethics derive from an alternative or richer conception of human nature—one that understands people as being motivated by love, friendship, responsibility, and caring rather than solely by self-interest and fear. A responsibility-based ethic, or an ethic of care, does not reject all the assumptions about human nature that undergird a rights-based ethic. Instead it contextualizes them, and at a minimum, it credits people in relationships with finer motivations and qualities. Although each ethic comes from different original premises about human nature, they are ultimately reconcilable if we can maintain an ongoing dialogue regarding both of them.[34]

Finding bridges from our current ethic's foundation in personal autonomy to a care-based ethic is critical to our making a successful shift. Autonomy, the power of an individual to control her own life and death, is as much a cornerstone of a care-based ethic as it is of modern medical ethics and legal practice.[35] The differences are in the sources and meanings of autonomy. In a care-based ethic, individual autonomy is a *process* nurtured in webs of relationships and responsibilities instead of a static condition pre-existing them.[36] Whereas the ideological basis of a rights-based ethic rests on an assumption of equally empowered, independent people, an ethic of care recognizes that many relationships contain dependencies between differently em-

---

additional works in feminist ethics on which I rely are NEL NODDINGS, CARING (1984); NEL NODDINGS, WOMEN AND EVIL 130-42 (1989); SARA RUDDICK, MATERNAL THINKING: TOWARD A POLITICS OF PEACE (1989); Annette C. Baier, *The Need For More Than Justice, in* SCIENCE, MORALITY & FEMINIST THEORY 41 (Marsha Hanen & Kai Nielsen eds., 1987) [hereinafter FEMINIST THEORY]; Lorraine Code, *Second Persons, in* FEMINIST THEORY 357; Ann Ferguson, *A Feminist Aspect Theory of the Self, in* FEMINIST THEORY 339; Marilyn Friedman, *Beyond Caring: The De-Moralization of Gender, in* FEMINIST THEORY 87; Virginia Held, *Non-Contractual Society: A Feminist View, in* FEMINIST THEORY 111; Alison Jaggar, *Feminist Ethics: Projects, Problems, Prospects, in* FEMINIST ETHICS 78 (Claudia Card ed., 1991); Carol S. Robb, *A Framework for Feminist Ethics, in* WOMEN'S CONSCIOUSNESS, WOMEN'S CONSCIENCE: A READER IN FEMINIST ETHICS 211 (Barbara Hilkert Andolsen et al. eds., 1985) [hereinafter WOMEN'S CONSCIOUSNESS]; Ruth L. Smith, *Feminism and the Moral Subject, in* WOMEN'S CONSCIOUSNESS 235; Joan C. Tronto, *Women and Caring: What Can Feminists Learn About Morality from Caring? in* GENDER/BODY/KNOWLEDGE: FEMINIST RECONSTRUCTIONS OF BEING AND KNOWING 172 (Alison M. Jaggar & Susan R. Bordo eds., 1989); Virginia Warren, *Feminist Directions in Medical Ethics,* 4 HYPATIA 73 (1989); Caroline Whitbeck, *A Different Reality: Feminist Ontology, in* BEYOND DOMINATION (Carol Gould ed., 1983); WHO CARES: THEORY, RESEARCH, AND EDUCATIONAL IMPLICATIONS OF THE ETHIC OF CARE (Mary M. Brabeck ed., 1989).

34. *See, e.g.*, Robin Dillon, *Care and Respect, in* EXPLORATIONS IN FEMINIST ETHICS: THEORY AND PRACTICE 69 (Eve Browning Cole & Susan Coutrap-McQuin eds., 1992).

35. *See, e.g.*, PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH: DECIDING TO FOREGO LIFE-SUSTAINING TREATMENT 26-27, 44 (1983); TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 67-119, 210 (3d ed., 1989).

36. *See, e.g.*, Smith, *supra* note 33, at 235.

powered people—parents and children, caregivers and mentally or physically impaired people, teachers and students, doctors and patients, and at times lovers and friends.[37] The autonomy of an ethic of care can be melded with the autonomy concerns in a rights-based medical ethic, if it is understood to mean self-governing moral agency, rather than independent or self-contained decision-making. Self-governing in an ethic of care does not mean governing alone by abstract reasoning and distant observations, but means choosing options with respect to responsibilities, relationships, conversations, and dialogues with others.

Autonomy, the premiere value in contemporary medical ethics, is transformed from a notion of independent decision-making to an interactive process of developing agency and empowerment through relationships, connections, and interdependencies. Caregiving becomes a means of empowering a cared-for person—of enhancing her autonomy. An ethic of care framework implores a caregiver to use his or her power, expertise, knowledge, and attention to respond and enable the cared-for person to communicate and meet her needs. If the empowerment to act as a moral agent or decision maker in one's own life is dependent upon the care or assistance of others, non-interference or failure to assist may be contrary to, rather than consistent with, autonomy.[38] In a care-based ethic, refusing care or assistance in particular contexts might be neglectful and unethical rather than obedient to abstract norms.

We need to reconceptualize the physician's role as medical caregiver in light of an ethic of care in the context of our contemporary society, which has pursued technology and science to its outer limits. If we define the physician's relationship to a dying person as "giving medical care" rather than as prolonging life or healing, we need to redefine "giving medical care" as responding to the dying person's needs during the dying process. Legal and ethical questions about appropriate medical practice should be about "how best to care for" the person in need. Sometimes dying people have needs for radical technological interventions, sometimes for maintenance care, sometimes for pain relief and comfort, sometimes for security and dignity, and sometimes for aid in their dying process. All of those may be appropriate ways for doctors to give medical care, but the ethical propriety of a particularized method of caring is context-specific.

Caring for dying people requires careful attention to their particularized needs. The caregivers must discover what those needs are by listening to the patient; conversing with her and those who know her

---

37. Baier, *supra* note 33, at 53-56; Marilyn Friedman, *Feminism and Modern Friendship: Dislocating the Community, in* EXPLORATIONS IN FEMINIST ETHICS: THEORY AND PRACTICE 89 (Eve Browning Cole & Susan Coultrap-McQuin eds., 1992); Held, *supra* note 33; RUDDICK, MATERNAL THINKING, *supra* note 33.

38. Baier, *supra* note 33.

best and are responsible for her care; and learning about her options, beliefs, and her concerns for her well-being and the well-being of others about whom she cares. Depending upon the person and the context, these needs may be met by empowering the dying person to act for herself—whether by refusing potentially life-extending treatments, by utilizing self-administered, pain relief pumps, or by giving a prescription for a potentially lethal dose of drugs as Dr. Quill did. There will be times under a care-based paradigm where the giving of medical care by a physician is the giving of treatment that completes the dying process rather than elongates it. If this medical care responds to a patient's request for assistance in dying with dignity—a request which has been made after ongoing conversations with family, friends, and caregivers that carefully considered all options—it is the ethical response of a physician to use her special knowledge and skills to help her patient implement this meaningful decision.

In shifting from a rule-based ethic to a care-based ethic we can also reclaim the dying process from a totally medicalized definition. By reclaiming it, rehumanizing it, and returning it to the person dying and the people with whom that person is interconnected, we can establish more agency, more responsibility, and more control over our own deaths. We can reclaim it as a process that centers on our bodies, but is about our lives, our roles, our relationships, and our connections.

Dying, particularly dying from illness or old age rather than from a sudden accident, is not a process involving only one person. Although the process focuses on the dying person's wants and needs, it is interactive, relational, and connected. It is social and communal. We show our love and care as a community when we act responsively and compassionately in accord with the dying person's needs. These are not abstract questions about isolated individuals. These are concrete processes in lives of interconnected people. Dying must be reconceived as the social, communal process it is. Decision-making about dying ought to grow out of ongoing conversations among interrelated people.

Participating in and responding to the dying person's experiences and needs is the caring response, the role of the physicians and health care workers, and the compassionate act. The doctor becomes one of a community of people involved in the process. She can share information, explain options, and implement treatment decisions made by patients with loved ones. The decision to end life ideally would be worked out collaboratively with multiple inputs, including the physician's, but it is not the physician's decision to make. The physician's role is to provide the requested medical care or to enable the patient to receive it.

Usually at this point, a doctor responds: "Why doctors? If you want people to aid others in ending their lives, why not let families do it or hire special people as executioners?" "We do not want the responsibility," say the doctors. "It is not our job." While I understand these arguments, I would respond to physicians that it is your

job. Part of a doctor's medical expertise is caring for dying people. This model does not empower doctors to make the decisions for patients; they are asked only to help implement patients' decisions. We are not increasing their responsibility beyond what they undertook when they agreed to provide medical care for a patient. We are legally empowering them to use their medical training and expertise to care for someone dying in a manner that is most compatible with their expressed needs. If this is what they were licensed to do, why should they remove themselves?

Families and friends, while in closer relationships with the dying person to help decide about appropriate avenues of care consistent with that person's needs, lack the necessary medical expertise and access to means of easy pain relief, or quick "death with dignity," to perform direct acts of assistance in dying. Lay people often have to resort to violence and crude methodologies, like guns and strangulation, to end someone's life. Even if given access to the drugs, they are unfamiliar with their administration and dosages, with what to do if difficulties arise, and with mechanisms for determining their success.

In addition, I would be remiss if I did not acknowledge the added emotional torment to a dying person of having to ask a friend or family member to assist in her dying and the emotional strain that such assistance must place on the loved one asked. It seems to me that the doctor's slight removal from the inner web of relationships puts him or her in a better position to give the medical care that ends life, if that is what the patient needs. I emphasize again that even though the process of dying is not a medical process, the physical action of giving life-ending medical care is.

IV. Conclusion

The important change that results from applying a care-based paradigm is the understanding of requested life-ending treatment as one form of medical caregiving for dying patients. We can establish guidelines that assure that patients are clear and consistent in their request and that they have discussed their decisions with friends, family, and caregivers. The guidelines should not be an impediment to implementing a person's end-of-life option for medical assistance but a mechanism for preventing abuse. Under a care-based analysis, the option of physician assistance may give dying people the security, dignity, and control that Dr. Quill spoke of giving Diane. That would be empowering and consistent with autonomy.

In summary, my arguments are addressed toward cases like the one presented by Dr. Quill—terminally ill patients who request physician assistance to end their suffering during their dying process. At a minimum, the law and medical ethics must be able to respond appropriately to this easy case before it can tackle the more difficult ones. I have reserved for another day questions about terminally ill

versus non-terminally ill patients and questions about physical versus psychic pain. I explicitly avoided cases of patients unable to communicate their desires. I would ultimately hope that my arguments will serve to enrich conversations about those patients as well. I also focused on the role of the physician rather than other health care providers, in particular nurses, who play a critical role in caring for dying patients. Time limitations prevent me from addressing issues of nurses' roles here.

The crux of my argument is that we ought to alter the paradigm and language of our discussion about physicians' roles in care for the dying. By utilizing a care-based ethic, we can better realize goals of patient autonomy and dignity while emphasizing values of care, compassion, and responsibility.

Our battles over physician-assisted death seem to be smokescreens for our unwillingness to accept the inevitability of death. Our denial of death and the strength of the medical model to resist it at all costs have led to heroics, to violent interventions, and to prohibitions against acting in furtherance of dying people's needs when those needs are to die. If we use feminist ethics to reconceive of death as a process of dying in particularized people's lives and we come to understand the role of medicine as caring for rather than prolonging life, where caring can include multiple ways of responding to dying peoples' needs, our legal system can make spaces in its laws to legitimize rather than punish or wink at that kind of compassionate, caring medical response.

SELECTED BIBLIOGRAPHY

Abel, Emily K., & Margaret K. Nelson, CIRCLES OF CARE: WORK & IDENTITY IN WOMEN'S LIVES (SUNY Press, 1990).

Abrams, Frederick R. *Letter—Physician Participation in Assisted Suicide*, 263 JAMA 1197 (1990).

Ames, Katrine, et al. *Last Rights*, NEWSWEEK, Aug. 26, 1991, at 40.

Andolsen, Barbara Hilkert, Christine E. Gudorf & Mary D. Pellauer, eds. WOMEN'S CONSCIOUSNESS/WOMEN'S CONSCIENCE: A READER IN FEMINIST ETHICS (Harper & Row, 1985).

Anonymous. *It's Over, Debbie*, 259 JAMA 272 (1988).

Baier, Annette C. *The Need for More than Justice*, in SCIENCE, MORALITY & FEMINIST THEORY 41 (Marsha Hanen & Kai Nielsen eds., U. Calgary Press, 1987).

Bailey, Don V. THE CHALLENGE OF EUTHANASIA: AN ANNOTATED BIBLIOGRAPHY ON EUTHANASIA AND RELATED SUBJECTS (University Press, 1990).

Baird, Robert M. & Stuart E. Rosenbaum, eds. EUTHANASIA: THE MORAL ISSUES (Prometheus Books, 1989).

Bartlett, Elizabeth Ann. *Beyond Either/Or: Justice and Care in the Ethics of Albert Camus* in EXPLOITATIONS IN FEMINIST ETHICS: THEORY AND PRACTICE 82 (Eve Browning Cole & Susan Coultrap-McQuin eds., Ind. U. Press, 1992).

Begley, Sharon & Mark Starr. *Choosing Death*, NEWSWEEK, Aug. 26, 1991, at 42.

Belkin, Lisa. *Doctors Debate Helping the Terminally Ill Die*, N.Y. TIMES, May 24, 1989, at A1.

Berger, Arthur S. & Joyce Berger, eds. TO DIE OR NOT TO DIE? CROSS-DISCIPLINARY, CULTURAL AND LEGAL PERSPECTIVES ON THE RIGHT TO CHOOSE DEATH (Praeger, 1990).

Bordy, Baruch A., ed. SUICIDE AND EUTHANASIA: HISTORICAL AND CONTEMPORARY THEMES (Kluwer Academic Press, 1989).

Callahan, Daniel. *Can We Return Death to Disease?*, Special Supplement, HASTINGS CENTER REP. 4-6 (Jan.-Feb. 1989).

Campbell, Robert, & Diane Collinson. ENDING LIVES (Basil Blackwell, Open University, 1988).

Capron, A.M. *Reflections on Health Law and Ethics*, 18 L. MED. & HEALTH CARE 15 (1990).

Card, Claudia, ed. FEMINIST ETHICS (U. Kan. 1991).

Cassel, Christine, & Diane Meier. *Morals and Moralism in the Debate Over Euthanasia and Assisted Suicide*, 323 NEW. ENG. J. MED. 750 (1990).

Casswell, Donald G. *Rejecting Criminal Liability for Life-Shortening Palliative Care*, 6 J. CONTEMP. HEALTH L. & POL'Y 127 (1990).

Cole, Eve Browning & Susan Coultrap-McQuin, eds. EXPLORATIONS IN FEMINIST ETHICS: THEORY AND PRACTICE (Ind. U. Press, 1992).

Cornell, Drucilla. BEYOND ACCOMMODATION: ETHICAL FEMINISM, DECONSTRUCTION AND THE LAW (Routledge, 1991).

Devettere, Raymond J. *Neocortical Death and Human Death*, 18 L. MED. & HEALTH CARE 96 (1990).

Dillon, Robin S. *Care and Respect*, in EXPLORATIONS IN FEMINIST ETHICS: THEORY AND PRACTICE 69 (Eve Browning Cole & Susan Coultrap-McQuin eds., Ind. U. Press, 1992).

Downing, A.B., & Barbara Smoker, eds. VOLUNTARY EUTHANASIA: EXPERTS DEBATE THE RIGHT TO DIE (Peter Owen Ltd., 1986).

Engelhardt, H. Tristram, Jr. *Fashioning an Ethic for Life and Death in a Post-Modern Society*, Special Supplement, HASTINGS CENTER REP. 7-9 (Jan.-Feb. 1989).

Engelhardt, H. Tristram, Jr. & Michele Malloy. *Suicide and Assisting Suicide: A Critique of Legal Sanctions*, 36 SW L.J. 1003 (1982).

Farley, Margaret A. *Feminist Theology and Bioethics*, in WOMEN'S CONSCIOUSNESS/WOMEN'S CONSCIENCE 285 (Barbara Hilkert Andolsen, Christine E. Gudorf & Mary D. Pellauer eds., Harper & Row, 1985).

Fisher, Berenice, & Joan Tronto. *Toward a Feminist Theory of Caring*, in CIRCLES OF CARE: WORK & IDENTITY IN WOMEN'S LIVES 35 (Emily K. Abel & Margaret K. Nelson eds., SUNY Press, 1990).

Fletcher, Joseph. MORALS AND MEDICINE 172-210 (Beacon Press, 1960).

Foot, Philippa. *Euthanasia* in ETHICAL ISSUES RELATING TO LIFE & DEATH 14 (John Ladd ed., Oxford U. Press, 1979) (originally published in 6 PHIL. & PUB. AFFAIRS 85 (1977)).

Frazer, Elizabeth, Jennifer Hornsby & Sabina Lovibond, eds. ETHICS: A FEMINIST READER (Blackwell, 1992).

Friedman, Marilyn. *Beyond Caring: The De-Moralization of Gender*, in SCIENCE, MORALITY & FEMINIST THEORY 87 (Marsha Hanen & Kai Nielsen eds., 1987).

Fry, Sara T. *The Role of Caring in a Theory of Nursing Ethics*, 4 HYPATIA 88 (1989).

Garbesi, George C. *The Law of Assisted Suicide*, 3 ISSUES L. & MED. 93 (1987).

Gaylin, Willard, et al. *Doctors Must Not Kill* in EUTHANASIA: THE MORAL ISSUES 25 (Robert Baird & Stuart Rosenbaum eds., Prometheus Books, 1989) (originally published in 259 JAMA 2139 (1988)).

Gibson, Count D. *Perimortal Initiatives: Issues in Foregoing Life-Sustaining Treatment, Suicide, and Assisted Suicide*, 3 ISSUES L. & MED. 29 (1987).

Goldberg, Steven. *The Changing Face of Death: Computers, Consciousness, and Nancy Cruzan*, 43 Stan. L. Rev. 659 (1991).

Goodwin, James S. *Mercy Killing: Mercy for Whom?*, 265 JAMA 326 (1991).

Gorney, Cynthia. *Cryonics and Suicide: Avoiding the 'Slippery Slope,'* WASH. POST, May 1, 1990, at D6.

Hanen, Marsha, & Kai Nielsen, eds., SCIENCE, MORALITY & FEMINIST THEORY (Canadian J. of Phil., Vol. 13, U. Calgary Press, 1987).

Horgan, John. *Science and the Citizen: Death with Dignity*, 264 SCI. AM. 17 (1991).

Humphry, Derek, & Ann Wickett. THE RIGHT TO DIE: UNDERSTANDING EUTHANASIA (Harper & Row, 1986).

Humphry, Derek. COMPASSIONATE CRIMES, BROKEN TABOOS (Hemlock Society, 1986).

Humphry, Derek. FINAL EXIT: THE PRACTICALITIES OF SELF-DELIVERANCE AND ASSISTED SUICIDE FOR THE DYING (Hemlock Society/ Carol, 1991).

Hunter, Samuel. *Active Euthanasia Violates Fundamental Principles,* 262 JAMA 3074 (1989).

Ikuta, Sandra Segal. *Dying at the Right Time: A Critical Legal Theory Approach to Timing-of-Death Issues*, 5 ISSUES L. & MED. 3 (1989).

Institute of Medical Ethics Working Party on the Ethics of Prolonging Life and Assisting Death. *Assisted Death*, 336 LANCET 610 (1990).

Kass, Leon R. *Neither For Love Nor Money: Why Doctors Must Not Kill*, 94 PUB. INTEREST 25 (1989).

Kantor, Jay E. MEDICAL ETHICS FOR PHYSICIANS-IN-TRAINING 111 (Plenum Medical, 1989).

Kennedy, Ian. TREAT ME RIGHT: ESSAYS IN MEDICAL LAW AND ETHICS (Clarendon Press, 1988).

Kevorkian, Jack. PRESCRIPTION: MEDICINE: THE GOODNESS OF PLANNED DEATH (Prometheus Books, 1991).

King, Patricia A. *The Authority of Families to Make Medical Decisions for Incompetent Patients after the Cruzan Decision*, 19 L. MED. & HEALTH CARE 76 (1991).

Kittay, Eve Feder & Diane Meyers, eds. WOMEN AND MORAL THEORY (Rowman & Littlefield, 1987).

Kuhse, Helga. THE SANCTITY-OF-LIFE DOCTRINE IN MEDICINE: A CRITIQUE (Clarendon Press, 1987).

Kuhse, Helga. *The Case for Active Voluntary Euthanasia*, 14 L. MED. & HEALTH CARE 145 (1986).

Lace, Timothy J. *The Physician Can Play a Positive Role in Euthanasia*, 262 JAMA 3075 (1989).

Ladd, John, ed. ETHICAL ISSUES RELATING TO LIFE & DEATH (Oxford U. Press, 1979).

Lifton, Robert Jay. THE NAZI DOCTORS: MEDICAL KILLING AND THE PSYCHOLOGY OF GENOCIDE (Basic Books, 1986).

Lind, Alice. *Hospitals and Hospices: Feminist Decisions About Care for the Dying*, in HEALING TECHNOLOGY: FEMINIST PERSPECTIVES 263 (Kathryn Strother Ratcliff et al. eds., U. Michigan, 1989).

Longmore, Paul K. *Elizabeth Bouvia, Assisted Suicide and Social Prejudice*, 3 ISSUES L. & MED. 141 (1987).

Meisel, Alan. THE RIGHT TO DIE (1989).

Miles, Steven H. & Allison August. *Courts, Gender and "The Right to Die,"* 18 L. MED. & HEALTH CARE 85 (1990).

Model Aid-in-Dying Act, 75 IOWA L. REV. 125 (1989).

Molenda, Francis. *Active Euthanasia: Can It Be Justified?*, 24 TULSA L. REV. 165 (1988).

Noddings, Nel. WOMEN AND EVIL 130-152 (U. Cal., 1989).

Noddings, Nel. CARING: A FEMININE APPROACH TO ETHICS AND MORAL EDUCATION (U. Cal., 1984).

O'Brien, Shari. *Facilitating Euthanatic, Rational Suicide: Help Me Go Gentle Into That Good Night*, 31 ST. LOUIS U. L.J. 655 (1987).

Orentlicher, David. *Physician Participation in Assisted Suicide*, 262 JAMA 1844 (1989).

Parachini, Allan. *The California Humane and Dignified Death Initiative*, Special Supplement, HASTING CENTER REP. 10-2 (1989).

Pellegrino, Edmund D. *Ethics*, 261 JAMA 2843 (1989).

Potts, Stephen G. *Looking for the Exit Door: Killing and Caring in Modern Medicine*, 25 HOUS. L. REV. 493 (1988).

Quill, Timothy. *Death and Dignity,* 324 NEW ENG. J. MED. 691 (1991).

Rachels, James. THE END OF LIFE: EUTHANASIA AND MORALITY (Oxford U. Press, 1986).

Rachels, James. *Active and Passive Euthanasia* in EUTHANASIA: THE MORAL ISSUES 45 (Robert Baird & Stuart Rosenbaum eds., Prometheus Books, 1989) (originally published in 292 NEW ENG. J. MED. 78-80 (1975)).

Ratcliff, Kathryn Strother, ed. HEALING TECHNOLOGY: FEMINIST PERSPECTIVES (U. Mich. Press, 1989).

Rhoden, Nancy K. *Litigating Life and Death*, 102 HARV. L. REV. 375 (1988).

Richman, Joseph. *Sanctioned Assisting Suicide: Impact on Family Relations*, 3 ISSUES L. & MED. 53 (1987).

Robb, Carol S., *A Framework for Feminist Ethics*, in WOMEN'S CONSCIOUSNESS/WOMEN'S CONSCIENCE 211 (Barbara Hilkert Andolsen, Christine E. Gudorf & Mary D. Pellauer eds., Harper & Row, 1985).

Robinson, Jane. *Euthanasia: The Collision of Theory and Practice*, 18 L. MED. & HEALTH CARE 105 (1990).

Rosenbaum, Ron. *Angel of Death: The Trial of the Suicide Doctor*, VANITY FAIR 147 (May 1991).

Rosenblum, Victor G. & Clarke D. Forsythe. *The Right to Assisted Suicide: Protection of Autonomy or an Open Door to Social Killing?*, 6 ISSUES L. & MED. 3 (1990).

Shaffer, Catherine. Note, *Criminal Liability for Assisting Suicide*, 86 COLUM. L. REV. 348 (1986).

Sherman, Rorie. *Bioethics Debate*, NAT'L L.J., May 13, 1991, at 1.

Sherwin, Susan. *Feminist and Medical Ethics: Two Different Approaches to Contextual Ethics*, 4 HYPATIA 57 (1989).

Singer, Peter A. *Should Doctors Kill Patients?*, 138 CAN. MED. ASS'N J. 1000 (1988).

Singer, Peter A. & Mark Siegler. *Euthanasia—A Critique*, 322 NEW ENG. J. MED. 1881 (1990).

Skegg, P.D.G. LAW, ETHICS, AND MEDICINE: STUDIES IN MEDICAL LAW 121-139 (Clarendon Press, 1984).

Smith, George P. II. *All's Well That Ends Well: Toward a Policy of Assisted Rational Suicide or Merely Enlightened Self-Determination?*, 22 U.C. DAVIS L. REV. 275 (1989).

Smith, Ruth L. *Feminism and the Moral Subject*, in WOMEN'S CONSCIOUSNESS/WOMEN'S CONSCIENCE 285 (Barbara Hilkert Andolsen, Christine E. Gudorf & Mary D. Pellauer eds., Harper & Row, 1985).

Special Issue, *Feminist Ethics and Medicine*, 4 HYPATIA No. 2 (1989).

Sprung, Charles L. *Changing Attitudes and Practices in Forgoing Life-Sustaining Treatments*, 263 JAMA 2211 (1990).

Steinbrook, Robert. *Support Grows for Euthanasia; Medical Killing Runs Against Thousands of Years of Professional Tradition, But Interest Is Increasing as Advances Allow Doctors to Prolong Life and Sometimes, Patients' Suffering*, L.A. TIMES, Apr. 19, 1991, at A1.

Taylor, Rodney L. & Jean Watson, eds. THEY SHALL NOT HURT: HUMAN SUFFERING AND HUMAN CARING (Colo. Assoc. Univ. Press, 1989).

Thomson, Judith Jarvis. *Killing, Letting Die, and The Trolley Problem* in RIGHTS, RESTITUTION AND RISK 78 (William Parent ed., Harv. U. Press, 1986).

Tronto, Joan C. *Women and Caring: What Can Feminists Learn About Morality from Caring?*, in GENDER/BODY/KNOWLEDGE: FEMINIST RECONSTRUCTIONS OF BEING AND KNOWING 172 (Alison M. Jaggar & Susan R. Bordo eds., Rutgers U. Press, 1989).

Vaux, Kenneth. *Debbie's Dying: Mercy Killing and the Good Death*, in EUTHANASIA: THE MORAL ISSUES 29 (Robert Baird & Stuart Rosenbaum eds., Prometheus Books, 1989) (originally published in 259 JAMA 2141 (1988)).

Veatch, Robert M. A THEORY OF MEDICAL ETHICS (Basic Books, 1981).

Wainey, Deborah A. Note, *Active Voluntary Euthanasia: The Ultimate Act of Care for the Dying*, 37 CLEV. ST. L. REV. 645 (1989).

Wanzer, Sidney H. et al. *Special Article: The Physician's Responsibility Toward Hopelessly Ill Patients*, 320 NEW ENG. J. MED. 844 (1989).

Warren, Virginia. *Feminist Directions in Medical Ethics*, 4 HYPATIA 73 (1989).

Watson, Jean. *Human Caring and Suffering: A Subjective Model for Health Sciences*, in THEY SHALL NOT HURT: HUMAN SUFFERING AND HUMAN CARING (Rodney Taylor & Jean Watson eds., Colo. Assoc. Univ. Press, 1989).

The Sound of Silence Breaking: Catholic Women, Abortion, and the Law

TERESA GODWIN PHELPS*

How does one speak about abortion as a Catholic feminist legal scholar? What language can one find that balances the critical values involved in the abortion controversy without either mindlessly mouthing the position taken by the Church hierarchy or sounding hopelessly secular? Must one choose between a secular language that seems not to value potential life nor recognize the moral issues involved in any abortion and an abstract religious language that seems not to value women?

These questions began to trouble me as I found myself more and more alienated from the extreme positions espoused by the so-called "pro-life" and "pro-choice" proponents. Because I had no answers to them, I remained silent on the issue of abortion. Unable to find a vocabulary, I, like so many women on so many issues, simply did not speak.

This article is about my search for answers to the above questions and my quest for a voice and vocabulary in which to ask them. It begins first, as I did, with a survey[1] of some of the material written by Catholic scholars on the topic of abortion. It turns, as I discovered I needed to, to conversations with Catholic women, some of whom I had known and worked with for years without ever discussing abortion. These women, like me, had been silent (or silenced) and many of them were fearful of raising questions in the shadow of a Church that claimed to have all the answers. This article does not pretend to resolve the pressing emotional and moral issues to which

---

* Professor of Law, University of Notre Dame Law School; B.A., M.A., Ph.D., Notre Dame; M.S.L., Yale.

1. In this article I do not try to provide an in-depth survey of writings on abortion, even of those writings done by Catholics. The volume of writing on abortion is truly monumental, and I have tried to indicate where readers may go for other surveys. The purpose of this section of the article is to give an overview of the breadth of opinion coming from Catholic writers. Other relevant material not discussed here includes MARY ANN GLENDON, ABORTION AND DIVORCE IN WESTERN LAW: AMERICAN FAILURES, EUROPEAN CHALLENGES (1987); HANS LOTSTRA, ABORTION: THE CATHOLIC DEBATE IN AMERICA (1985); Michael W. McConnell, *How Not To Promote Serious Deliberation About Abortion,* 58 U. CHI. L. REV. 1181 (1991) (reviewing Laurence Tribe, *Abortion: The Clash of Absolutes* (1991); James R. Kelly, *Abortion: What Americans* Really *Think and the Catholic Challenge,* 165 AMERICA (No. 13) Nov. 2, 1991.

abortion gives rise. Instead, it is the story of conversations, conversations that open the way for fruitful dialogue.

THE CATHOLIC WRITINGS

Catholic theologians, legal scholars, and others writing about abortion offer a confusing cacophony of opinions, often complete with historical evidence in support of directly conflicting positions. The extreme ends of the spectrum of opinions might be represented by the work of John Noonan, on the one hand, and Barbara Ferraro and Patricia Hussey on the other. John Noonan's book, *A Private Choice: Abortion in America in the Seventies*,[2] appears at first blush to be promising. The title alone seems to indicate that this early work in response to the legalization of abortion recognizes the intimate and private nature of an abortion decision. The initial chapter establishes that "[t]he 'issue' of abortion is not a single dispute over a hairline distinction. It is many-faceted. It has turned into a multiplicity of issues."[3] Noonan, however, uses the title ironically. He argues that "each act of abortion bears on the structure of marriage and the family, the role and duties of parents, the limitations of the paternal part in procreation, and the virtues that characterize a mother."[4] He also maintains that abortion is anything but a private choice and that "[i]ndividual federal judges have not hesitated to set aside the oldest of laws—that on murder itself . . . ."[5] It is unfortunate that Noonan shows the rigidity of his position so early in the book because he alienates readers who might more willingly follow his painstaking tracking of the historical condemnation of abortion as well as his intelligent and thought provoking discussion of the notion of liberty in western political philosophy. The abortion cases, he concludes, fly in the face of American legal history and tradition. He also concludes the "abortion liberty" (as the book comes to call it) is not "compatible with our Constitution, our family structure, our notion of governmental power, and our sense of the human person."[6] He further states "[t]here must be a limit to a liberty so mistaken in its foundations, so far reaching in its malignant consequences, and so deadly in its exercise."[7] Noonan's account is persuasive and moving as he compels the reader to face the "hard reality"[8] of abortion.

---

2. JOHN T. NOONAN, JR., A PRIVATE CHOICE: ABORTION IN AMERICA IN THE SEVENTIES (1979).
3. *Id.* at 2.
4. *Id.* at 3.
5. *Id.* at 1.
6. *Id.* at 4.
7. *Id.* at 192.
8. *Id.* at 4.

The more recent *No Turning Back: Two Nuns' Battle with the Vatican Over Women's Rights to Choose*[9] similarly traces the history of abortion and concludes abortion traditionally has been *accepted*, not condemned. Hussey and Ferraro's "hard reality" is grounded in their daily work with poor women and men at their drop-in center in West Virginia. They gathered data as they struggled to clarify for themselves their own position on abortion after the Catholic Church censored them for signing (with twenty-two other nuns, four priests and sixty-nine Catholic lay people) a full-page advertisement in the *New York Times*[10] that claimed Catholics held a diversity of opinions on abortion. This advertisement responded to the attacks launched by conservative Catholic bishops against vice-presidential candidate Geraldine Ferraro for her pro-choice position on abortion. The Vatican quickly reacted by demanding the nuns and priests either recant or face dismissal from their religious orders.

During the time Ferraro and Hussey were negotiating their position with their religious orders and the Vatican, they spent a year researching historical and theological data on abortion. As they uncovered instance after instance of the Church's misogyny and oppression of both sexuality and women, Hussey and Ferraro's position was transformed from one of asking for the opening of dialogue on abortion to a full-fledged pro-choice stand.[11] They, like Noonan, were and are life-long Catholics. Like Noonan, they are well-educated—albeit in theology rather than in law. Their work, like his, is moving and persuasive.

More promising for me because of its openness is the work of Catholic priest and theologian Richard McCormick, S.J., and the work of Daniel and Sidney Callahan, a husband and wife who differ on abortion and frequently present their arguments in a dialogic fashion. McCormick argues for elevating the level of discourse to a true dialogue instead of perpetuating the "dialogue of the deaf"[12] consisting of opposing monologues droning on and people refusing to listen to each other. He exhibits a sensitivity to the complexity of the abortion issue, writing:

> Abortion is a matter that is morally problematic, pastorally delicate, legislatively thorny, constitutionally insecure, ecumenically divisive, medically normless, humanly anguishing, racially provocative, journalistically abused, personally biased, and widely performed. It

---

9. BARBARA FERRARO, PATRICIA HUSSEY WITH JANE O'REILLY, NO TURNING BACK: TWO NUNS' BATTLE WITH THE VATICAN OVER WOMEN'S RIGHT TO CHOOSE (1990).

10. N.Y. TIMES, October 24, 1984.

11. *Id.* at 261.

12. RICHARD A. MCCORMICK, S.J., HOW BRAVE A NEW WORLD?: DILEMMAS IN BIOETHICS 176 (1981). *See also*, Richard A. McCormick, *Abortion: The Unexplored Middle Ground*, 10 SECOND OPINION: HEALTH, FAITH, AND ETHICS 41 (1989).

> demands a most extraordinary discipline of moral thought, one that is penetrating without being impenetrable, humanly compassionate without being morally compromising, legally realistic without being legally positivistic, instructed by cognate disciplines without being determined by them, informed by tradition without being enslaved by it, etc. Abortion, therefore, is a severe testing ground for moral reflection.[13]

McCormick attempts to increase the possibility of real dialogue by laying down some rules for discussion and by establishing that "it is the right and the duty of conscientious citizens to continue to debate this matter in the public forum."[14] Among his rules for discussion is the necessity of incorporating women's perspectives in the abortion discussion, although he objects to the frequently heard claim that a celibate man, or any man, should not speak about abortion. He writes: "The more one knows experientially of a situation, the more sensitive one ought to be to the situation's many-faceted circumstances, though I believe this is frequently not the way things turn out. Self-involved agents are frequently self-interested agents with a one-dimensional view of things obvious to most reasonable and reflective people."[15] McCormick makes one further point that I want to raise here because I think my conversations shed some light on it: "What Americans as a culture think about sexuality and how they live it will have a strong influence on their evaluation of fetal life and abortion."[16]

Among the women McCormick mentions as having spoken on the abortion issue is Sidney Callahan, an associate professor of psychology. Callahan describes herself as a pro-life feminist,[17] and she confronts the typical pro-choice arguments with what she calls a "different perspective on what is good for women."[18] She compares the dependent, disempowered fetus to the historical condition of women when they were seen, both legally and religiously, as incorporated into the "one flesh" of their husbands and as having a lesser or no soul. She argues that pro-choice feminists overvalue individual autonomy and should instead emphasize an expanded sense of responsibility. She also sees pro-choice feminism as too easily

---

13. McCormick, How BRAVE A NEW WORLD, *supra* note 12, at 118-19. This chapter, titled "The Abortion Dossier," provides a detailed account of positions taken by various Catholic theologians and critiques of their rationales. It also includes an appraisal of the papal position as well as an analysis of positions taken by various groups of bishops.

14. *Id.* at 174.

15. *Id.* at 201.

16. *Id.* at 202.

17. Sidney Callahan, *Abortion and the Sexual Agenda*, COMMONWEAL, April 26, 1986, at 232.

18. *Id.* at 234.

embracing a male-oriented, permissive, and erotic view of sexuality that is foreign to the sexual nature of most women.

Sidney Callahan's husband, Daniel Callahan, a philosopher and director of the Hastings Center, takes a different view. He would leave the abortion choice to women, but "[o]nce women had the choice, it would become important for them in their private lives to give thought to what would count as a morally justifiable choice; and it would be no less appropriate to have some public discussion about the standards and criteria appropriate for such choices . . . ."[19] He laments that the pro-choice movement has not been open to discussing the moral dimension of abortion and states "its inability or unwillingness to come to grips with the moral issue threatens its political credibility."[20]

After her fieldwork among activists on both sides of the issue, Faye Ginsburg unexpectedly concludes there is a crucial need for dialogue.[21] She becomes, in the end, neither pro-choice nor pro-life, but pro-dialogue. She quotes a woman speaking about a group of opposing women who did get together to discuss the issue: "the stereotypes are disappearing as we work together toward a common goal."[22] Ruth Colker reaches a similar conclusion in her writings about abortion. She exhorts that "[f]eminists need to be more open to discussion about the *process* used to reach various conclusions,"[23] and that they should enter into dialogue not to persuade others but to listen and to empathize.[24]

Unfortunately, the Church's[25] official response to any opening of the discussion on abortion has been swift and sure, as promised by its response to the 1984 *New York Times* advertisement. Archbishop

---

19. Daniel Callahan, *An Ethical Challenge to Prochoice Advocates: Abortion and the Pluralistic Proposition,* COMMONWEAL Nov. 23, 1990, at 681. The Callahans have also edited a collection of essays on abortion titled ABORTION: UNDERSTANDING DIFFERENCES (1984). In their introduction to this volume they explain its impetus in their ongoing discussion and disagreement about abortion that has gone on for nearly twenty years: "And yet we still disagree. How can it be, we ask ourselves, that intelligent people of goodwill who know all the same facts and all the same arguments still come down on different sides of the controversy? . . . Our curiosity, not only about why we differ, but about why others differ as well, was the impulse behind this book." *Id.* at xv.

20. *Id.* at 682.

21. FAYE D. GINSBURG, CONTESTED LIVES: THE ABORTION DEBATE IN AN AMERICAN COMMUNITY 225 (1989).

22. *Id.*

23. Ruth Colker, *Feminism, Theology, and Abortion: Toward Love, Compassion, and Wisdom,* 77 CAL. L. REV. 1011, 1036 (1989).

24. *Id.* at 1033.

25. I am using Church with a capital *C* throughout this article to stand for the Roman Catholic Church. Nonetheless, I understand such usage is controversial because there are other meanings for both catholic and church.

Rembert Weakland of Milwaukee instituted postabortion counseling for women in his archdiocese and held six "listening sessions," in which Catholic women were invited to air their views on the Church's stand on abortion. Weakland summarized what he heard in articles in the archdiocesan newspaper: the women saw abortion as a tragedy but also deplored the tactics and narrowness of pro-life organizations. They also did not support the Church's ban on contraception.[26] Weakland was censured and disempowered for holding these sessions and for writing about them.

This article is a reaction to that silencing and a response to what I hear as a call to dialogue. It also represents a personal need as I searched for a discourse about abortion that did not give way to abstraction, that was dominated neither by theologians nor legal scholars but could inform both disciplines.

The final section of the article attempts to connect the conversations to legal policies. Although the conclusions reached are mine alone, they are rooted in and inspired by the conversations.

THE CONVERSATIONS

Although I guaranteed anonymity to each woman with whom I spoke, I can say some general things about them. With one exception, all the women whose remarks are related in this article describe themselves as Catholic. Some responded with a resounding "Yes!", perhaps conveying a sense of resoluteness in the face of their differences with the official Church. Others responded "Yes," but with something self-deferential, such as "Rome might not consider me one." The ages of the women range from the mid-twenties to sixty. They are both white women and women of color, married and unmarried women, mothers and nonmothers. Some had been members of religious communities. My selection, though broad-based, was not random; rather, it was controlled by my perceptions of the women. Although I had no prior knowledge of their individual positions on abortion, I did select women whom I knew took themselves seriously as spiritual persons, women who acknowledged and lived in accordance with God's presence in the world. I knew each would speak her own mind, influenced by little outside of her own moral deliberations. I also knew these deliberations were neither shallow nor trivial, nor were they secular. These conversations are with serious Catholic women who have reached their positions on abortion after much reflection and personal experience. These differ, I think, from many feminist narratives because I make no claim of typicality. The mere fact these women are Catholic makes them

---

26. *See generally*, Paul Wilkes, *Profiles: The Education of an Archbishop I*, THE NEW YORKER, July 15, 1991, at 49-53.

atypical, and even among Catholic women these women might not be considered typical in that they are professional women, many with advanced degrees. I relate what they said to me to break a long, harmful, and painful silence, to bring forth voices that can be of significant value to others of us grappling with the problem.[27]

Interviewee A had been a member of a religious community. She said she arrived at a pro-choice position after lengthy reflection and personal experience. For a long time, her position was that taken by many Catholic politicians: although she was personally against abortion and would not have one herself, she did not believe it should be illegal. Lately, however, her position had shifted. During a pregnancy, she had reason to fear for the health of the fetus. She decided to undergo tests that would determine whether or not the fetus was deformed. If the tests indicated deformity, she and her husband had agreed they would abort the fetus. Just before the test, she miscarried. Realizing she was in fact capable of having an abortion caused her to modify her initial position. She said she continued to fear abuses, but we should start with the right and then worry about abuses, such as sex selection and abortion used as birth control. Abuses are those practices that result from taking life lightly, a life-is-not-sacred attitude. "What I came to understand during my own pregnancy," she said, "is that any decision about the fetus had to be mine. That, above all, women must have the right to choose for themselves."

"The Church," she continued, "makes broad decisions based on the greater good. It does not consider exceptions and makes decisions with possible abuses in mind. I felt, during my pregnancy, that I was an exception, that there was no greater good in giving birth to a severely deformed baby. The Church does not guarantee that it will care for this child if anything happens to me. Who would care for this child was my greatest fear. Sin is turning away from love. Is it loving to bring a child with severe handicaps into the world? I don't think so."

Would you allow or require any controls on the abortion decision?

---

27. The interview proceeded as follows: I called each woman and told her I was writing a law review article on abortion, and I wanted to talk with her about her position on abortion and how she reached it. I said I would be taking notes and publishing the conversations but I would not identify the women with whom I spoke. Every woman I called agreed to talk with me. I began each interview by asking the woman to talk about her position on abortion. I then asked (in nearly every case) four standard questions—more to keep the conversation going than to acquire specific information. These questions are indicated in italics. I took careful notes and did not edit except for redundancy.

What I found most interesting and touching is the fact that when I thanked the women for talking with me, they thanked me, saying they were happy their views would somehow, albeit anonymously, be made known. They were grateful because I had listened to them.

> I'd like to assure more thoughtful decision-making. I wouldn't object to a twenty-four hour waiting period. The father should be involved in the decision, and I have little problem with parental consent. Anything that would encourage a woman to look for possible sources of support. I suppose I have a bias in favor of giving birth, if possible, if it is what the woman decides.

What would you do if you were counseling a woman considering an abortion?

> I would try to discover exactly where the woman was, what kind of connection she had with herself. I would counsel without bias, or with a bias only toward the woman's own autonomy. I think that pro-choice can also be pro-life in a broader sense of the term.

What is the status of the fetus? When does life begin?

> The fetus is life, maybe not fully human. I only know for certain that human life begins at some point during pregnancy and that I would know when that was. I would have a feeling of connection to this life.

Do you feel tension in being a Catholic and your position on abortion?

> Yes.

Interviewee B focused less on a personal epiphany and more on her experience with sexist attitudes within the Catholic Church, attitudes that denied women moral autonomy. She saw abortion as a necessary evil. It is necessary in that not everyone should or can follow through in giving birth; legal or not, women have always and will always have abortions. It is evil because destruction is not a good and the fetus is life, perhaps even human life. She takes this position because she believes we must create a situation in society in which women are their own moral agents and what happens to them is exclusively the choice of the women.

Would you allow or require any controls on the abortion decision?

> No. There should be no controls. In an ideal world I would discourage abortion for purposes of birth control and after viability. But we do not live in an ideal world, one in which a woman would receive one hundred percent support for any choice she made.

What would you do if you were counseling a woman considering an abortion?

> I would insist on *talk*, on conversation about the decision. I would ask her to consider how she would feel afterwards. I would ask her to consider all her options. I would be fully neutral. It is not my place to judge.

What is the status of the fetus? When does life begin?

> This is the most troubling aspect for me. The fetus is life, human life. But sometimes human life has to be killed; there is a conflict of rights. Abortion is, in this way, an evil, but a necessary evil.

Do you feel tension in being a Catholic and your position on abortion?

> I believe that the hierarchy of the Church is *wrong* and not representative of the real Church as defined by Vatican II, which is the people. For the Church hierarchy, women are not moral agents, they are mere carriers of babies. The hierarchy takes an absolutist position, favoring the fetus over the woman, without saying why.

Interviewee C similarly sees the need for women to be their own moral agents, but she put the situation differently. Although she believes the abortion debate is controlled by extremists on both sides and many abortions are immoral, she said criminalizing abortion is a greater evil. She described her position as "muddy," saying it reflected her background. She was a Catholic girl who attended Catholic elementary school and later CCD (Catholic Christian Doctrine) classes. While she was in high school, the Catholic Youth Organization asked her to give a presentation on abortion arguing abortion was linked to premarital sex. She began to think the fight against abortion was a front for the Church's effort to have social (particularly sexual) control over teenagers. She then began to wonder why there were so many abortions, why there were so many unwanted pregnancies, why women could not protect themselves better.

Her position was also affected by her knowledge of history. It was not until the 1840s that men began to say "abortion is murder." Knowing this made her critical of men's authoritarian claims.

Would you allow or require controls on the abortion decision?

> Abortion as birth control is immoral, but there's no way to control it. Government should not be regulating reproduction. It would do better to find ways to teach women to respect themselves and each other. It should involve itself in education about sexual responsibility. Abortion is a matter of conscience and the morality of abortion is a province of the Church not the government.

What would you do if you were counseling a woman considering abortion?

> I would tell her that whatever she did to act fast. The age of the fetus makes a critical difference, although I would not draw the line for her. I'd have no bias toward her having the child. I would ask her what she thought the fetus was and encourage her to act with that in mind. Therefore, if, for her, human life began at conception, an abortion would be an immoral act. The question is: What is our responsibility in our own particular situation?

Do you feel tension in being a Catholic and your position on abortion?

> No. In that the Church's position on abortion is an effort to protect women, it is noble. The problem is that the Church puts the onus of what it is to be caring on women alone. I am more concerned about the situation of women in society. There is far too much pressure on women to be sexually available to men. Prostitution is so fully integrated into our society that it's invisible. The image of woman is that of sexual object—of access for men. The self gets sold. Rather than putting so much emphasis on abortion, the Church should concern itself with helping people (men and women, boys and girls) learn to handle their own sexuality, on education about sexual responsibility. I'd like to see a movement that addresses why there are so many unwanted pregnancies. I would quit my job and devote full time to that if I could figure out a way to do it. The Church should help people discover what makes them happy. How do people come to know themselves? Like themselves? Then there would be little need for abortion.

When I asked Interviewee D if she would describe herself as Catholic, she answered, "Absolutely!" She immediately went on to say her view of "Catholic" was larger than the Pope's. She had a post-Vatican II view of the Catholic Church that defined it as the people of God rather than the hierarchy.

On abortion, she said women are either moral agents or they are not. If they are moral agents, then they have to have a choice. The Church and state are not the moral agents in the case of an abortion and thus cannot make the requisite moral decision. All sex should not result in childbearing. That is a biologistic view of women as mere carriers of babies.

Would you allow or require any controls on the abortion decision?

> Although women are moral agents, they should consult with others; thus I wouldn't object to a twenty-four hour waiting period. I would want her to get the best advice possible, but that is not tied to a parent, spouse, or the father. I don't want as many abortions as there are, so I think that we (as women, as people, as friends) can be with people and help them—either with the decision or with the child.
>
> Abortion should definitely be allowed in cases of rape, incest, or danger to the health of the mother. If I knew I was carrying a deformed fetus, I might think about an abortion. It's hard for me to sort out how I do (or should) feel about motherhood. So much about mothering is socially constructed. I would prohibit abortion after viability; this is the only place where I think the law should have a role.

What would you do if you were counseling a woman considering an abortion?

> I personally know no one who has had an abortion (that I know about) and in counseling I don't know what I would do. What bothers me, though, is that I feel as though I can't talk about

> abortion at all. The women here [at Notre Dame] don't have the benefit of what I know, or the opportunity to talk with someone like me.

What is the status of the fetus? When does life begin?

> I don't know and it may be that abortion is a killing. Yet we say in the Catholic tradition that all killing is not morally wrong. Why can't there be a "just abortion" ethic as there is a "just war" ethic? The ethicists who could help us with this are silenced or written off as not "real" Catholics.

Do you feel tension in being a Catholic and your position on abortion?

> No. The hierarchical Church is wrong. The Pope's position on birth control is hypocritical; birth control is a moral good. The Church has not held a consistent position on abortion through time, and we have changed our understanding of when life begins. As for the Church argument that abortion exploits women, the oppression of women is so complex that that argument oversimplifies the issue.

Interviewee E described herself as a Catholic "on the margin," a "questing Catholic" reflectively examining theological views. She would not criminalize any abortions. She said her own experience had changed her mind about abortion. First, in studying *Humanae Vitae*, she discovered the Pope appointed people to study the issue of birth control and then ignored their recommendations. This made her distrust the Church position. Second, when she was in her forties, after she had been told not to have any more children (she already had several), she thought she was pregnant. Although it turned out she was not pregnant, she realized she probably would have aborted the fetus. If she could have an abortion, then she could not continue thinking it immoral.

Would you allow or require any controls on the abortion decision?

> No, although ideally all abortions should occur in the first trimester. There are evils: sex selection and birth control, but still they should not be prohibited legally.
>
> I'm not sure about parental consent, but I would not require spouse's or father's consent. I would, though, insist on some discussion about the decision.

What would you do if you were counseling a woman considering an abortion?

> A woman should not make a choice in isolation; she should discuss her situation with her mate, a doctor, a priest, a counselor, anyone whom she trusts, and explore her options. I would try not to influence her either way. In fact, I think that there would be fewer abortions if women felt that real, unbiased, unjudgmental counseling was available to them. As it is, many young Catholic women panic, tell no one, and get an abortion. If they had someone to turn to

> who would not judge them, they might find a support system in place that would enable them to have the baby. I would go so far as to accompany a woman if she decided to have an abortion. And afterwards, she could turn to me for forgiveness.

What is the status of the fetus? When does life begin?

> The fetus is not a human person until the third trimester. I have had miscarriages and did not feel as though I had lost a baby, but a potential baby.

Do you feel tension in being a Catholic and your position on abortion?

> No. The Church's position is a moral wrong and the Church has lost its credibility, especially on issues concerning women. The Church is particularly wrong in pressuring Catholic politicians into pro-life stands. It is just a single Catholic voice talking about abortion in the United States. Other Catholic voices must be heard, but how can it be? The only role the Church should have in a woman's decision is that of support, counseling, and forgiveness. The Church should have no role in influencing the government's control of abortion.

Interviewee F described herself as a Catholic. She said she is undecided on abortion, but that she does not totally embrace the Church's position. Principled but individual circumstances made her see the need for some abortions. For example, when a friend's daughter, who was dangerously out of control on drugs, became pregnant, Interviewee F said she could see how an abortion might be a wise choice. She thinks the "just war" theory should apply in cases when a woman's life is in jeopardy and should be considered if she had been victimized by an unjust aggressor (as in rape). "The right to life stand appalls me," she said, "in its monolithic position and its lack of regard for human beings already living. On the other hand, the pro-choice position does not encompass the moral seriousness of the decision."

Would you allow or require any controls on the abortion decision?

> I would not support a law that would outlaw abortion under all conditions. I would like to see a more liberal law than that. I don't have the right to enforce a doubtful moral area on someone else. I would require some waiting period during which a woman would be required to get some counseling on what issues were involved for her and what alternatives were available to her (although I think it is probably impractical since such counseling would be difficult to monitor). For me abortion is not just a woman's issue; it is broader than that. The fetus is still a product of two people and the father should have some input into the decision.

What would you do if you were counseling a woman considering an abortion?

> I would try to be as open as possible and help her explore her options. In fact, I have been in such a counseling situation and told the young woman that I could not have an abortion under those circumstances but that I would continue to support *her* as a person even if I couldn't support her choice. I don't think that I would accompany a woman to have an abortion, though. I might give her money that she could use for taxi fare. I know this sounds hypocritical, but it clearly reflects my own ambivalence about abortion. I don't like it, I don't want to be complicitous in it, yet I would not want to see it made criminal. And I would not judge a woman who decided to have an abortion.

Do you feel tension in being a Catholic and your position on abortion?

> Yes and no. The Church's stand on birth control gravely weakens its moral authority on abortion. Its position on birth control is based on false, out-dated science rooted in suspicion of women's sexuality. But it's not just a plot to keep women down, it's much more complicated than that. I see the Church's position as based in a fear of women and their procreative power which must be kept in control. It's tied to a sense of property and of passing on the male line. Abortion is a complex moral issue that the Church does not address fully. And, sadly, to some extent I am fearful of speaking about it.

Interviewee G, who described herself as a Catholic, said her position was similar to New York governor Mario Cuomo's position: she personally could not abort, but she does not believe abortion should be illegal. She knows she could not abort because she had been in circumstances when an unwanted pregnancy might have been possible, and she had confronted the issue of what she would do if she became pregnant. Her position on the legality of abortion is based on the "lawyer in her" and on her respect for other viewpoints that treat abortion as a morally acceptable decision. She was thus pro-choice. The fact that even religions are split on this question as to what is morally acceptable had been very influential on her.

She took a different view of the moral autonomy of women than many other interviewees. She feels there is too much emphasis on the individual at the cost of community values. She explained her view of the interaction of autonomy and community in this way:

> Personally, under *no* circumstances would I have an abortion. I believe that a level of surrender is needed in life in order to be able to grow. But in order to surrender, you need to have some moral autonomy. Otherwise, the pregnancy is just another kind of blow that buffets a woman about. So it is really the morally autonomous woman who can decide *not* to abort. A woman needs a sense of personal power in order to surrender to life and to larger community values.

Would you allow or require any controls on the abortion decision?

I am not happy with the thought of late pregnancy abortions or with abortion for gender selection. But since the state cannot distinguish a woman's reasons for having an abortion, it should stay out of it.

Any consent requirement presupposes we live in an ideal society in which everyone treats everyone else with respect. But we don't. A minor should consult with some adult but not necessarily a parent. Some conversation should take place.

As an aside, I find the exceptions that people are willing to make for cases of rape and incest very troubling. It seems to focus on whether or not a woman *decided* to have sex and then punish her for it.

What is the status of the fetus? When does life begin?

There is much uncertainty about this, about "life," about "soul." We can never know when soul is infused. For me, I resolve that doubt by saying that although I couldn't have an abortion, I can't answer that question for everyone.

Do you feel tension in being a Catholic and your position on abortion?

I feel no tension with the Church. I've disagreed with the Church on issues since I was about 6; disagreeing with the Church has always been part of my life. The Pope and the Church hierarchy mean little to me. But I am not just a "cultural Catholic." I have found a community of people with whom I pray. It is Catholic and so am I.

The Church should be talking to its children much differently about sex, including birth control. Sex is immoral for everyone when it is not "creative" or "life-giving," but that doesn't just mean conceiving a child. Also the Church should be a community for people in helping them make decisions and standing with them. There is a word in Spanish that means something like "accompaniment" that captures what I think the Church's presence in our life decisions should be.

The Church and the law are in a position to and do control women's lives. If, on the other hand, the Church and the law helped to empower women, they would be more likely to keep their babies. Surrender, acceptance, has to come from a person who feels whole.

Interviewee H was raised a Catholic but became uncomfortable identifying herself as a Catholic because she could not abide by certain Church tenets. She described herself as religious rather than a secular humanist. "The crucial thing," she said, "is to focus on spirituality." She described herself as a supporter of women's options in life as opposed to the constraints from external perceptions of what women should be. "There should be no legal prohibitions on abortion. When I think about abortion, I think about a woman who is in a position she doesn't want to be in. The ideal is that women get pregnant only when they want to and are qualified to be mothers.

Abortion is unfortunate but not wrong. It is an indication of a larger social problem: immature, adolescent sexuality in society. By this I mean that even people past adolescence have an immature, adolescent attitude toward sex, not that the problem of abortion is confined to adolescents. The incidence of abortion could be reduced if we addressed underlying social problems. There is, of course, a moral dimension to the decision. Morally, it would be better if unmarried women who are not ready for parenthood did not get pregnant. People should think more seriously about sexual relations."

Would you allow or require any controls on the abortion decision?

> No, not parent, spouse, or father. Ideally, it would be all right but the way that it actually works out is that it is an impediment to abortion and increases the suffering of the woman. The only value in a consent law is that a woman is not alone in making the decision.

What would you do if you were counseling a woman considering an abortion?

> If it were an immature girl, my bias would be toward her having an abortion. If it were a young career woman, I'd be in favor of her having the child if she wanted to. I would not counsel anyone to have a child without first ascertaining her inclination. I would support an affirmative inclination or a negative inclination because a child should be born into circumstances where she or he is wanted. The difference is that the potential child will not suffer by being born and the potential mother will not suffer because of the birth. When a woman is counseled to have a child, not enough attention is paid to the consequences. This should be the only affirmative role for the state: to help a woman with the consequences of having given birth.

What is the status of the fetus? When does life begin?

> The fetus is life. The egg and sperm before conception are life. We define human life according to how we want people to treat an object. Human life for me occurs at birth when human personality begins to manifest.

Do you feel any tension with your spirituality and your position on abortion?

> No. But it is troubling. The important value for me is that no injury occur and abortion is some kind of injury to potential human life and perhaps to others.

Interviewee I said she considered herself a Catholic, although she was not sure the Church would consider her one. She said legally she was pro-choice, but she was not sure she would ever have an abortion herself. A woman will know if an abortion is a moral choice for her and it should be left up to her.

Would you allow or require any controls on the abortion decision?

> I would not allow abortions after viability. The government should regulate doctors, make sure that they are licensed so that they don't hurt women. That's what it already does but will cease to do if abortion becomes illegal.
> Minors should not make the decision alone so parental consent laws are all right. No spousal or paternal consent, though.

What would you do if you were counseling a woman considering an abortion?

> I wouldn't put myself in that position. I would send her to someone else, someone qualified to counsel her.

What is the status of the fetus? When does life begin?

> I'm not sure that the fetus is a human being. Since I'm not sure, I can't impose my beliefs on someone else. We can draw a line, though, at viability. Life definitely begins at viability.

Do you feel tension in being a Catholic and your position on abortion?

> No. I've never accepted all the Church's teaching. What the Church says is of no concern to me.

Interviewee J described herself as a Catholic. She is pro-choice. She feels women should be empowered to control their own bodies. She is, however, not pro-abortion; abortion has to be a well-thought-out moral and conscious decision. She has held this position since high school and it has been affirmed by both her intuition and her observations of life, fairness, and how unexpected childbearing could inhibit a woman's potential.

Would you allow or require any controls on the abortion decision?

> No. The government's role should be normal licensing of doctors and of providing federal funds for poor women. Some abortions can be immoral—when used as a form of birth control or as revenge in a divorce situation—but the law should impose no restrictions. It's a moral choice that the woman needs to make herself, weighing outcomes and looking at her life situation.

What would you do if you were counseling a woman considering an abortion?

> I would encourage her to discuss her situation with as many people as she felt she could. There is a need for discussion before an abortion; suffering occurs when the decision is not reflected on. I would have no bias either way.

What is the status of the fetus? When does life begin?

> It is potential life. The killing of potential life occurs with abortion.

Do you feel tension in being a Catholic and your position on abortion?

> Yes. I feel alienated from the Church. But I have no problem calling myself a Catholic. I see myself as a new breed of Catholic. The Church's role should be guidance on moral issues; it should be open to dialogue, forgiveness, and acceptance. The number of abortions could be reduced by means of birth control education and with more wisdom is selecting sexual partners.

I have duplicated these conversations so individual readers may form their own conclusions and may draw from them what they find useful. Nonetheless, there are certain general themes I would like to underscore. First, the women differ across the board on their personal moral positions on abortion. They range from having decided never, under any circumstances, to have an abortion to actually having decided to have one under certain conditions. None, however, placed her personal moral claims on other women. Second, they differ as to when they believe life begins, although this belief was not necessarily linked to their positions on abortion. Some who believe life begins at conception still do not see abortion as necessarily immoral. Third, although some of the women are personally opposed to abortion and see it as generally immoral, none wants a complete legal prohibition. Instead, most see the abortion problem as societal, with its causes rooted more in the disempowerment and oppression of women. Fourth, there is a clear strain of thought regarding the necessity of dialogue about both an individual abortion decision and about abortion in general. Fifth, they share a general consensus about the need for women to have and to take responsibility over their own bodies and for their own sexuality. Sixth, their relationship with the Church hierarchy demonstrates a general disaffection for and alienation from policies concerning women and sexuality.

Before turning to what relevance these conclusions might have for the law, I would like to return to some of the Catholic writings. Despite my general agreement with much of what Richard McCormick writes, these conversations seem to belie his concern over the bias of women as self-interested agents having a one-dimensional view.[28] The legality or illegality of abortion has no personal relevance to some of the women with whom I spoke; they would not have abortions under any circumstances. What they bring experientially to the abortion question is serious, possibility-based reflection. Such

---

28. Michael McConnell seems to understand this point. In his review of Tribe's book, he writes that "[w]omen's attitudes toward abortion, both for and against, are often more complex. Even those who favor abortion rights often understand the act as the taking of human life, and feel grief, pain, and responsibility for it." McConnell, *supra* note 1, at 1191. He contrasts this to Laurence Tribe's "antiseptic and abstract treatment of the phenomenon." *Id.*

reflection is of unquestionable benefit because it comes from those who have dealt with or might actually have to deal with an unexpected pregnancy, those for whom there are no external legal constraints. For them abortion is neither illegal nor categorically immoral. The silencing or diminishing of the fruits of such reflection is an incalculable loss.

These conversations tend to disprove another point that McCormick makes: "What Americans as a culture think about sexuality and how they live it will have a strong influence on their evaluation of fetal life and abortion."[29] Although seriously stated by McCormick, the same point is often raised more crudely: Easy sex (promiscuity) requires easy access to abortion. Arguments for keeping abortion legal are thus dismissed as being products of a sexually permissive culture. These conversations reveal quite contrary thinking, behavior and reflection. If the positions of most of these women fit broadly into the category of "pro-choice feminism" opposed by Sidney Callahan, they are far from embracing a male-oriented, permissive, erotic view of sexuality. Rather, all their perspectives are decidedly woman-oriented and nonpermissive, espousing a responsible view of sexuality.

IMPLICATIONS FOR THE LAW

In a recent article outlining the criticisms often leveled against the use of feminist narratives in legal scholarship, Kathryn Abrams wrote:

> Establishing previously unheard perspectives as credible accounts of a social problem is the first step in feminist narrative persuasion. But it is also necessary for feminist scholars to convince their readers that these perspectives can contribute to legal change. This is, of course, another area in which some critics of feminist narrative have expressed doubts. These critics argue that narrative accounts do not clearly implicate particular legal rules or choices; that authors do not make clear the way in which narrative descriptions translate into normative proposals; and that the normative proposals suggested by narrative scholars are insufficiently developed to provide guidance to legal actors.[30]

Although Abrams does not fully credit this criticism, she does allow that the lack of legal prescriptions coming from legal scholars creates difficulties for readers looking for the same relationships between narrative and legal prescription as between doctrinal analysis and legal prescription. That relationship, she explains, is not so direct or

---

29. *See supra* notes 12-13.

30. Kathryn Abrams, *Hearing the Call of Stories*, 79 CAL. L. REV. 971, 1030 (1991).

necessarily immediate. Among the first of feminist narratives to emerge were those of "excluded voices," women "whose voices had not been heard in social discussion of a problem."[31] Because it seems to me the voices of Catholic women are in this category, among those voices hitherto unheard, I am hesitant to jump toward immediate remediation or to recommend specific legal rules. At the same time, I am reluctant not at least to enter the legal fray on these women's behalf.

First, the conversations underscore the paucity of the law's emphasis on rights. The current abortion debate is generally framed using rights language: the fetus or the woman has the superior right to life or to autonomy. A libertarian rights approach to abortion, the one most generally voiced by pro-choice activists, already has been revealed to fall short in cases in which economic factors are present. The rights approach

> assimilates disparate experience to a false Enlightenment universal. Not surprisingly, it captures most closely the experience of white, educated, middle-class women . . . . [F]or them, the libertarian emphasis on choice was consistent with experience. As a result, the needs of other particular groups of women—especially the poor—were 'pruned out' of the pro-choice discourse."[32]

In other words, poor women cannot choose whether to bear a child; their choice is compelled one way or another because of their economic situation.

Similarly, the language of rights fails to encompass the experience and beliefs of many of the women with whom I spoke. For many of them personally, choice had little meaning as long as abortion was a morally unacceptable alternative. And the raw handing to them of "choice" silences them as much as the dictates of the Vatican. It seemingly tells them they have what they need and they should seek no other remedy. This strikes me as demonstrably untrue and unfair. What does the law do for women who have no choice—economically, morally, or otherwise? What does the law do for women who choose or would choose to carry the pregnancy to term?

In a similar fashion, the language of guilt and sin circumscribes the choices of some Catholic women in ways the Church does not anticipate. If confessing the underlying "sin" of sexual intercourse

---

31. *Id.* at 1033.

32. Elizabeth Mensch and Alan Freeman, *The Politics of Virtue: Animals, Theology and Abortion,* 25 Ga. L. Rev. 923 (1991). This article provides an exhaustive analysis of theological approaches to abortion. It expands its critique of "choice" into a valuable discussion that is too rarely put forth by noting that there are numerous subtle pressures that might compel a woman toward an abortion: economics, feminist pressure to be autonomous and independent, male pressure to be sexually available. *Id.* at 1125.

outside of marriage is requisite to discussing how one should deal with a pregnancy, a woman might well decide not to discuss it at all. In so doing she may forgo conversations that would help her ascertain her own moral position and might help her understand how she will feel. Further, she may be blinded to the support system that might well enable her to have a child. Her choice is clouded because of her guilt about the pregnancy.

Thus far the law merely guarantees choice, but the guarantee is often empty of content. For only a few women does choice have any real meaning. Certain conditions precede choice just as certain conditions follow it. Choice has content only for empowered, employed, self-assured women who have been protected from nonconsensual sex, who have been educated about sexual responsibility and birth control, and who have lived in an environment free from sexual pressure from men (women who do not see their worthiness as linked to their sexual availability). Only when a woman's decision to bear a child is as fully supported by law (she will *not* lose her job, her seniority, her educational opportunities, her apartment, her place on the Honor Society, her good name) as her decision not to bear a child, does choice have any content.

Regardless of one's position on choice, the current law still fails. Under current law, women are left alone to make a crucial moral decision that will surely have an impact on their lives. This right to be left alone means, of course, that the state will not compel a woman to bear a child. She is thus far free from that sort of state intrusion into her life. The state, however, does not aid her in making a fully informed choice (state-supported abortion counseling may not be available). It remains to be seen whether the state may compel her to counsel with those whom she would not choose to counsel (a parent or husband, perhaps even the very person who forced sex on her). The state does not support her decision to abort if she does not have money by paying for the abortion, and the state does not support her decision to carry the pregnancy to term by guaranteeing she will not be economically or socially punished for becoming pregnant or for nurturing a child. Additionally, the state does little, and that inefficiently, to coerce the very people it may force her to consult when making a choice to support her financially, let alone emotionally, through her pregnancy and child-raising. When a woman says, as many do, "I have no choice," *this* is what she means. Just as it is misguided to speak of the fetus but not its mother as pro-life advocates do, it is also misguided to speak of the pregnant woman without the fetus as pro-choice advocates do. Pregnancy is a singular situation, completely without analogy to any other, in that the two, the woman and the fetus, are, in fact, one.[33] As long as

33. Pregnancy, in fact, seems to conceptualize the situation of which Kenneth

the mother or the fetus is seen as murdering or invading the "other," we are missing the point. When a woman says, "I have no choice," *this* is the situation in which she finds herself.

The legal abortion debate, much like the Church discussion of abortion, focuses on a very limited aspect of a woman's sexuality: her childbearing ability from the moment after conception to the moment after birth. What men and society in general do to a woman's sexuality, how she sees herself, how she comes to respect herself and her body, how she controls her life, how children are supported are, they tell us, beyond the reach of the law. This is not startling news, to be sure. Legal scholars like Robin West have been arguing for years that words on which the law seems fixated like "choice," "autonomy," and "power" are a poor fit for the values necessary for well-being in a woman's life.[34]

If, as we may conclude from these conversations, even some Catholic women, although they would not consider having abortions and although they do not know with certainty when life begins, still do not want abortion made illegal, does this mean we must abandon the law altogether in the abortion question? If it is true, as I believe it is, that rights language utterly fails to conceptualize the reality of pregnancy,[35] in which the pregnant woman and the fetus are interdependent, not autonomous individuals exercising rights against each other, must we abandon all legal language in discussing abortion? And if it is true, and I believe it is, that "law . . . tells stories about the culture that helped shape it and which in turn it helps to shape: stories about who we are, where we came from, and where we are going,"[36] what stories does our abortion law tell about our society?

I tread cautiously here. For the most part history reveals that the law has not been of much help to women, especially not when it enters the realm of pregnancy. Rather, women's potential or actual motherhood has traditionally been used against them and prevented women from entering professions and competing on an equal footing

---

Karst writes when he advocates that "our courts . . . need to look beyond the idea of rights as personal zones of noninterference to a conception of justice that recognize[s] our interdependence." Kenneth Karst, *Woman's Constitution*, 1984 DUKE L.J. 447, 471.

34. *See, e.g.*, Robin West, *Feminism, Critical Social Theory and Law*, 1989 U. CHIC. LEGAL FORUM 59.

35. The problem with rights language in another pregnancy-related issue, drug abuse by pregnant women, is delineated in Note, *Rethinking (M)otherhood: Feminist Theory and State Regulation of Pregnancy*, 103 HARV. L. REV. 1325 (1990). The author writes that "the prevailing conceptualization of the problem as a conflict between maternal and fetal rights is both illegitimate and counterproductive" and argues for a "rethinking of pregnancy based on a feminist understanding of connection and responsibility . . . ." at 1325-26, 1342.

36. MARY ANN GLENDON, ABORTION AND DIVORCE IN WESTERN LAW: AMERICAN FAILURES, EUROPEAN CHALLENGES 8 (1987).

with men. The laws regarding women's ability to bear children have traditionally limited their options. In numerous ways, the law has intervened negatively in constraining and controlling women's lives.

Yet, I am sure, it is possible the law might also intervene positively, in ways that help women, expand their opportunities, leave open their options, and support their decisions. In an effort to discern which laws offer positive intervention and which ones offer negative intervention, we need constantly to ask: Does this law (or how might it) promote and support women's autonomy and self-esteem?

We may turn for assistance in this crucial discernment process to an expanded version of a question raised earlier: What story does a particular law tell about us and about our society? What image does this law give us of the nature of women? Clearly restrictive abortion laws that take decision-making out of the hands of women[37] and place it under the power of the government tell a story about women and society that portrays women as less capable of moral reflection than the state. These laws do not encourage women to be autonomous and moral, but instead tell them they cannot be so trusted. What story does the law's requirement that a women receive consent from her parents or her spouse tell? In part, albeit in a skewed way, it echoes the sense from the conversations that a woman should discuss her decision with others. What does mandating who those others are mean? Especially if those others are those who have exercised control (perhaps perniciously) over the woman's life? It is true that many women see a crucial need for discussion before an abortion decision. The law, then, instead of telling a woman to whom she must speak (her father, mother, husband) might work toward supporting structures in which real conversations about an abortion decision may take place (by "real" I mean discussions that are not driven by bias on either side, with people who have no political stake in her decision).

What story does the law's ambivalence about women's sexuality, reflected in official and unofficial attitudes toward sexual harassment, date/acquaintance rape, marital rape, and other acts of sexual aggression toward women tell? It portrays women as needing to be sexually available to men. It buttresses the *sub rosa* message that a woman's worthiness is linked to her sexuality. What story does the continued use of constitutional protection to permit the manufacture and distribution of pornographic material that degrades women tell about society's regard for women's bodies and their sexuality? Do

---

37. Ruth Bader Ginsburg criticizes *Roe v. Wade* on a similar ground because the opinion makes a woman too dependent on the judgment of her doctor. *See*, Ruth Bader Ginsburg, *Some Thoughts on Autonomy and Equality in Relation to* Roe v. Wade, 63 N.C. L. REV. 375 (1985).

any of these things encourage women to be sexually responsible, to take control of their lives and their bodies? Do they encourage women's self-esteem? What story does our military establishment's (our government's) willingness to subject its servicewomen to insulting second class treatment even as they fought to defend Kuwait tell? Does it suggest strong support for women's equality? Or does it hint that such equality is fungible, controlled by circumstances?

One speaks as a Catholic woman about abortion by first of all *speaking*, by refusing to succumb to the comforting world of silence. Second, one refuses to accept the current state of the discourse as given. One does not surrender to the dichotomous language of rights but instead raises questions and subjects the rigid positions to insistent scrutiny. One insists that the law on abortion is a law affecting women and children and is thus interrelated to all other laws and practices that concern women and children.

We may never live in an ideal world and we may never agree on the morality of abortion or when human life begins. We can, nonetheless, work toward this world, an ideal world in which empowered women live in and are supported by their community. In such a world women are trusted to make their own choices, and both sides of the abortion debate, pro-life and pro-choice, might be surprised at what choices they make. We can begin this essential work by listening to each other.

Choice

MARY F. WHITE*

Two weeks after Dr. Chen confirmed her unplanned, unwanted pregnancy, Rachel was still waking up every morning thinking maybe today she would get her period. Of course by now, it wouldn't be her period, but a miscarriage. As the hours passed each day, hope fell and despair rose, making her sometimes irritable, sometimes depressed. Her two children, Alice and Daniel, like lightning rods absorbed her distress and then grounded themselves by fighting and crying. Kenneth, her husband, came home each hot summer evening to dreadful scenes of Rachel screaming at the children or, worse, Rachel in the bedroom with the door closed while the children (aged six and two) fought in the living room.

Rachel had no morning sickness and no fatigue. She continued to wear her ordinary clothes, which she buttoned even if it hurt. When she could, she pretended nothing was happening. Kenneth, engrossed in his own struggle for tenure and feeling guilty for having impregnated her, preferred not to bring up the subject. The only reminder between them of her condition was that the few times they made love they didn't bother to use any birth control. She thought about the baby constantly and yet insisted on living as though it didn't exist.

She had said nothing yet to the children, although when she was pregnant with Daniel she had told Alice right away, even though Alice was only three at the time. Now, trying to decide whether to get an abortion, Rachel kept the news of her pregnancy from her children.

"They're entitled to some protection from the troubles of the adult world," she explained to Fern, her best friend from college and a frequent visitor.

"I couldn't agree more," Fern said emphatically. Kenneth was out of town and Fern's husband Larry was having his friends over for their monthly poker game, so Fern had come over to Rachel's for dinner. Fern had been trying for years to have children but

---

* Director, Writing and Advocacy Program, University of Michigan Law School. B.A., 1970, Vanderbilt; J.D., 1973, University of Colorado. I wish to thank Margaret Marshall, Virginia Purvis-Smith, and Carol Yorkievitz for their helpful comments. This story is from a novel in progress titled *Rachel*. The following story, *Women in the Law*, is adapted from the draft of another novel titled *Untying the Knot*.

hadn't been able to conceive, and she considered herself an honorary aunt for Rachel's children. After dinner, she read stories to Alice and Daniel while Rachel cleaned up after the meal. Together they put the children to bed, and then the two women sat in the kitchen with glasses of iced tea and a bowl of popcorn.

"Do you think the fetus needs protection against me?" Rachel asked.

"I didn't mean that. Not consciously, anyway. I thought I was just agreeing with your decision not to say anything yet to Alice. The temptation to let something slip must be pretty great." She reached for a handful of popcorn. "On the other hand, maybe it is the fetus I'm thinking of. You know, if you did tell the kids you were pregnant, you'd have to have the baby. You could never explain an abortion to them. Doesn't that make you think you shouldn't do it? That you couldn't explain it to your children?"

"There are lots of things I can't explain to my children yet—rape, murder, nuclear war."

"See what I mean? I know I'm biased because of what Larry and I have been through trying to get pregnant, but what a list!" Fern said.

"That's not all though," Rachel insisted. "I don't talk to them about Kenneth's and my sex life, or about which of their friends I don't like, or about my mother's medical problems. I don't explain very much about my political opinions. Some things are private, and some things they're just not ready for. I'm sure I could tell them later, when they're older." Neither woman said anything for a minute or two. Rachel took a handful of popcorn and ate it piece by piece. Then she said, "Think how much you want to be pregnant. That's how much I don't want to be. You go through all kinds of procedures to try to get pregnant and everybody thinks that's fine, a good thing. I think about an abortion, and because I'm not a 16-year-old raped by her uncle, everybody thinks it's terrible. The worst thing is, even I think it's terrible. It's like there's an anti-abortion conspiracy, and I'm a member of it. For every one time I think it's my body and I'm entitled to control it (including by getting an abortion if that's what I think I have to do), there are five times I think, but it's a *baby*, and a woman is made to have babies. That's really it. I feel like the survival of the species is programmed into my genes and no matter how hard I try I can't escape it."

"Maybe it's better not to try."

"Fern," Rachel said, her voice rising to a scream, "how can you say that? You're the one who's supposed to be a feminist! Being able to escape their biology is what has made it possible for women to be something more than just wives and mothers. You've said a thousand times that the pill was the greatest invention of the twentieth century. You've campaigned for Medicare abortions and counter-demonstrated at Planned Parenthood!" She could have gone on, but

seeing the look of guilty misery on Fern's face she stopped.

"I know," Fern said in a small voice. "Maybe I was wrong. No, I don't mean that. In the abstract I'm all for a woman's right to an abortion. And poor women should have the same rights as rich women. Even in the concrete, I wouldn't disown you if you got one." Fern ran her fingers through her hair, dark and curly like Rachel's, but unlike Rachel's, cut so short it fit her head like a bathing cap. "Maybe if it were anybody but you I wouldn't feel this way. This sounds a little crazy, but I feel like you're having the babies I don't seem able to have. Alice and Daniel are very important to me. I'd love it if you had another baby." Fern's face was flushed and her brown eyes were shining with tears.

"Oh, Fern," Rachel said, starting to cry too. She was crying for Fern, the tragedy of her barrenness, but also for herself. How could Fern do this to her? You couldn't have a baby just because your friend wanted you to.

"I'm sorry, Rachel. I probably shouldn't have told you that."

"No," Rachel said, drying her eyes with her shirt, "I'm glad you did."

"Anyway, it's only my feelings. I'm not telling you what you should do. But still," Fern said, taking another handful of popcorn, "I don't really understand what the problem is. You're a good mother. You have a husband who'll support you. Why don't you want another, even if it isn't exactly planned? Thinking you can plan everything means spending an awful lot of time being disappointed. I speak from experience."

"I know, I know." Rachel rubbed at her face, then fixed an unruly lock of hair behind her ear. "It's not rational. Just something about the whole situation makes me feel so helpless and angry and awful. We have sex, right? And not because we're weird or bad or abnormal, but because that's what people do. And we use birth control because we're sensible, responsible people and we don't want to have a baby right now. And somehow, the birth control doesn't work and I end up pregnant. And then I have to choose between having a baby—a baby, mind you, not a new car or a new brand of toothpaste, but a lifetime relationship, and maybe even a lifetime responsibility because you know not all babies are born healthy and even when they are, things happen. I think about the stories I see in the paper about the seventy-year-old couples with forty-year-old children who can't take care of themselves . . ."

"That's not very likely."

"No, I suppose not, but it could happen. So that's on the one hand. My other choice is to have an abortion. When I asked my doctor to explain the procedure, she said it was a kind of birth, and my body would feel afterwards like it did after I had Alice and Daniel. In other words, it wouldn't be simple, and I'd feel like a murderer.

"Ten times a day I ask myself the same question you asked. I'm not a high school kid; I have a husband who loves me and will support me; having kids is what I've been doing. So why am I thinking for a minute about not having this baby? I don't know the answer, but sometimes I think what it comes down to is that I'm just afraid I can't do it this time. See, if I am a good mother, one reason is that so much of my energy goes into Alice and Daniel. Not that I hover over them all the time—I don't. At least I don't think I do. Yet, I don't really do anything else. It hasn't been true lately—the last few weeks I've been a mess, yelling at them or hiding from them—but usually I'm there for them if they need me. I mean for everything. I not only change Daniel's diapers, I keep track of when he's likely to need changing so he doesn't get a rash. When Alice has a friend over, I listen to how things are going so when voices start to rise in frustration, I can call them in for a snack or suggest a project. I plan my shopping in short spurts so I'll finish before Daniel gets exhausted and cranky. I talk to them and I play with them. And that's just the physical stuff. All the time I'm tuned to them, listening to what they're telling me in body language as well as words. All the time. And the hardest is when they're babies. Do you see what I'm saying?"

She didn't wait for an answer. "I imagine having this baby when Daniel isn't even three and my heart sinks. Alice has always been a sweetheart, but she was unbelievably demanding at that age, and she didn't have a new baby to cope with. Daniel will be even more difficult." Rachel took a deep breath and sighed. "And then," she said, "I remember very clearly what it was like the first few weeks after Daniel was born. Trying to make time for both children (not to mention for Kenneth or me) was excruciating. It's not even so easy now. If I have three of them looking up at me like little robins with their mouths open, I'm afraid I'll go crazy. Literally. What kind of mother can I be to any of them if that happens? Why should a baby who doesn't exist yet be such a trump card, especially over two children already here, alive, and needing what feels like 98% of what I've got to give?"

"I hate to say it, but I still don't get it. If you're saying the kids take too much out of you, why don't you just get more help?"

"Argh! Fern! Have you been listening to me at all? It's not just a question of getting more help—it's who I am and who my kids are to me. I'm afraid if I have this baby, it will spend its whole life feeling I didn't want it and I couldn't bear that. More help wouldn't do anything about that. Oh, God. I don't want to be thinking about this at all. I want not to be pregnant!"

"Well, I can see that all right, but Rachel, you can't run the movie backward. You're not in the situation of choosing whether or not to try to get pregnant—you have to choose whether to keep the baby that's already started."

Rachel sighed with hopeless resignation. "I want some more iced tea. You too?" She picked up the two glasses and went to the refrigerator. Over her shoulder she said, "I think I know when it happened. We had a weekend alone for the first time since before Daniel was born. This is the price we have to pay."

Fern waited till Rachel returned to the table. "Don't be ridiculous. These things happen. In fact, if you're not sterilized or on the pill, they happen pretty often. It could just as easily have happened the month before or the month after."

"I suppose I know what I'm going to do," Rachel continued, as if Fern hadn't spoken. "I'll torture myself with indecision for another few weeks until it's too late to do anything but have the baby, which is maybe what I want anyway. If I had an abortion I think I would feel it meant no more babies ever—psychologically speaking—and I'm not sure, not positive, that's what I want. I'm only thirty-three."

Fern patted Rachel's arm and in doing so noticed her watch. "Oh, no, look what time it is! I told Larry I'd be home by now!" Fern never checked the time without discovering—always to her great surprise—that it was a lot later than she'd thought. Fortunately for her students (she was an English professor) there was a large clock in every classroom.

"Okay," Rachel said, but neither of them got up. "I wish you lived closer, though I suppose I should be grateful it's Toledo rather than Tucson."

"Me too, but Larry wouldn't move to Ann Arbor unless I had a job here, and the University doesn't have any openings in my field. So," she shrugged, "I guess I stay where I am. Besides, I like my students. I'm happy where I am."

"Do you remember how we used to sit in the dorm in college and discuss our futures?" Rachel asked abruptly. "We assumed we'd have kids someday, but we'd also have important, fascinating, useful careers and there would be no conflict, no problems."

"And look at us now?" Fern questioned. "I have a career but no kids. You have kids but, at the moment, no career and we both have conflicts all over the place."

"Yeah. Neither the kids nor the career are what we expected."

"True, but so what? I wish I had kids, but I like what I'm doing and I think it's important. When I teach my students to criticize a novel, I'm also teaching them to criticize what they see on television or read in the newspaper. God knows that's necessary work. And there's plenty of meaningful work to do with the women students. They don't think, like our mothers did, that they have to stay home and keep house. In a way it's worse. They think all the battles are won, that women now have equal rights and there's nothing left to fight for. They think being a feminist means being frumpy and/or a lesbian, and they think there's nothing worse than that."

"So you're doing good in the world, but what about me? I'm just staying home and churning out babies. Forever, apparently, just like my mother."

"Oh, Rachel. What you're doing is the most important of all."

Rachel made a noise that started out as a contemptuous snort and turned into a laugh, then a cry. "Give me a break, Fern! You wouldn't do what I've been doing and you know it. If you ever manage to have a kid, you'll hire some other woman, likely Black or Hispanic and likely with kids of her own, to take care of yours. And you'll talk about how good she is with your child, but you'll pay her minimum wage and you won't really think what she does is as valuable as what you do. The people who are most sentimental about children are the ones who wouldn't spend a whole day with a child for a million dollars. It's hard work taking care of kids, which is why the thought of another one depresses me so much."

Fern looked stunned by her friend's explosion. "I'm not sure I know how to respond to all that. I shouldn't have to tell you some women don't have the choice to stay home with their kids. And not all the ones who could, should."

Rachel rubbed her forehead as if the pressure from her hands could relieve the internal pressure. "I'm sorry, Fern. I'm not being fair. I know perfectly well not every mother can stay with her kids. And I'm not at all interested in forcing women to stay home. I wish I could explain. You said before, why don't I just get more help. I can't afford a nanny, so that would mean daycare. This will probably sound crazy to you, but for me it would be as hard to put a newborn in daycare as to have the abortion. Maybe harder. Right now the baby is just a bunch of cells. I'm responsible for it, but mostly just to take care of myself, not drink alcohol, get enough sleep, that kind of thing. Once it's born, everything is completely different; it's a person, and one whose well-being I'm responsible for in a whole new way. The physical parts of that are relatively easy. Babies need to be fed and changed and pretty much anybody can do that, but they also need to be cared for—they need to have some one person paying attention to them, thinking they're important, something like the way I did with Alice and Daniel and the way I'm afraid I won't be able to do this time. And if I can't do it, maybe I shouldn't have the baby. Do you see?"

Fern shook her head. "I've really got to go. I don't know if I understand or not. I want a baby of my own so much, I can't think straight." She stood up and rummaged in her purse for her car keys, found them, and started for the door. Then she turned back and hugged Rachel. "If you decide you want the abortion and need somebody to go with you, let me know, okay?" Rachel nodded. Then she followed Fern to the door and watched until the taillights of her car disappeared.

Women in the Law

MARY F. WHITE

Hill, Ledbetter & French's office party was always held in the late afternoon of the Friday before Christmas. Only the attorneys and their spouses and the top executives of their largest client were invited, and normally it was a relatively sedate affair. No one had ever danced on top of a table, and because the party was held in one big room there was nowhere for illicit sex to take place. The expectation was that you would have a drink or two, eat a few hors d'oeuvres to keep the alcohol from making you too silly, and leave in time for an only slightly late dinner.

Joan Bennett, a senior associate at Hill, was getting ready to go to the party, arranging her desk and running through her mental list of things to do the next day, when a client called with a question about some interrogatories she had sent him. It was twenty minutes before she could get rid of him. When she left at 6:00 she seemed to be the last attorney in the office. Fortunately, the party was only an elevator ride away, and the shrimp bowl was still two-thirds full when she got off on the fifty-second floor. Everyone was at least one drink ahead of her, and though she was not an eager drinker, she felt the need for a scotch and soda. For the first time in the six years she'd been attending the firm Christmas party her husband, Neil, would not be present. They had recently separated and on occasions like this she felt like half of a couple. Though Neil had always embarrassed her by talking too loud and making bad jokes, at the moment worrying about what he might say or do seemed preferable to not having him there at all.

She got her drink and with a glass in her hand felt instantly more secure. She surveyed the room full of men and women in dark suits. She knew almost all of them except for a few of the spouses and had good working relations with most of the attorneys. She thought of them as friends, but as she looked at them now, there didn't seem to be anybody she wanted to talk to. Conscious of hunger, she walked over to the shrimp bowl. There was always a shrimp bowl at the Christmas party and it seemed, with its countless enormous shrimp, a symbol of the wealth of the firm. She ate one and had her mouth full with a second when she realized there was someone behind her.

"Ah hah, I knew I'd find you here!"

Joan turned and nodded, swallowed. "Hi, Phil. Were you looking for me?" Phil Crisman was a partner in his mid-fifties, not much over five feet tall, with a bad back and a reputation in the firm for

being brilliant. Joan seldom talked to him except at firm parties and she had never been able to understand what the reputation rested on. Her experience was that he was unspeakably boring and, in spite of his size, overbearing. He always seemed to seek her out at these parties and when he found her, often at the shrimp bowl, would talk without a break for twenty minutes, paying no attention to anything she said or even seeming to notice whether she was listening or not. And at the same time he would engage her in a kind of dance, leaning toward her, shuffling forward a few inches, causing Joan to back up wherever there was room. The first year he had backed her up to a wall; after that she made sure to avoid such dead ends. One year, by lucky accident, she had backed to within a few feet of the ladies' room and made her escape that way. His topic was always the same, military history. Had she said something once to make him think she was interested, or did he find the topic so fascinating he couldn't imagine anyone less engrossed in it than he was? She always smiled and tried to learn something. She'd been brought up to be polite to her elders and even if she hadn't been raised that way, she couldn't afford to alienate him this year. She was up for partner.

She listened to Phil drone on, smiling and nodding at appropriate places and wondering how many drinks he had had, for he seemed to be making less sense than usual. She finished her own drink and held up the glass.

He broke off his analysis of the importance of tanks in the desert in World War II and asked, "Shall I get you another? I need a refill myself. What were you having?"

"No thanks. You go ahead. If you don't mind, I think I'll get myself a little something more to eat." And she turned away, practically into the arms of Chip Lucas. At the worst stage of the break-up of her marriage, when she knew there was no future in it but couldn't imagine how to get out of it, Joan had gone for a drink after work with Chip, a partner in his late thirties, known to all of the women in the office, from secretaries to partners, as Pinch Lucas. In her despair, one thing had led to another and she had gone to bed with him. Apparently the word had instantly gone around the firm, because every man interested in "scoring" had dropped casually by her office, just to chat, within a week. Chip, too, had asked her out again, but she had turned them all down. None of them had seemed to take offense, but even though the incident had happened months ago, it was only in the last few weeks that she'd been comfortable running into Chip and the others in the coffee room. Now she moved aside to let him pass.

"No, no, don't run away. Don't I get to wish you Merry Christmas?" He put his arm around her and kissed her on the cheek.

"I can't stop you, can I?" She laughed to take the sting out of the words.

"I love this time of year. Where's your husband? Neil, isn't that his name? He usually makes his presence felt."

"He's not here. We're separated."

"Oh, gee, I knew that. I'm sorry." Chip still had his arm around her and he gave her a little squeeze, which felt to Joan more proprietary than sympathetic.

"You don't need to be sorry. I'm not."

"Ha, ha, ever the wit, my dear. I take it you were the dumpor, not the dumpee? It's a role that suits you better." He finally removed his arm and looked at her with a quizzical smile. "So, Joan, how are things? How is life treating you these days? I haven't seen you much lately."

"I've been pretty busy."

"Yes, I know. The firm can't let all those brains go to waste. How could you associates ever support us in the style to which we are accustomed if you didn't work night and day? Ha, ha. And are you enjoying it?"

"Well, I wouldn't mind getting more of my weekends off, but it's interesting. Why do you ask?"

"No reason. I just like to know if our associates are happy. Have you ever thought of going anyplace else?"

"No, why should I? I've always been happy here." With a sudden sinking feeling she asked, "Chip, are you just making conversation, or is there some ulterior motive here?"

"Joan, how your mind runs! Such a suspicious girl—I mean woman. Of course I'm just making conversation. But maybe we've run out of topics. Anyway, I see somebody I need to talk to. Merry Christmas, Joan." He raised his glass in a salute, then walked off toward the first year associate who was rumored to be his current conquest.

Oh, God, now what? Joan looked around the room. It was close to seven. A few people were starting to leave for dinner. If Neil had been here, they probably would have joined two other associates and their spouses for dinner in Chinatown as they had the last few years. The other couples had asked her if she wanted to join them but Joan had refused, thinking she would feel too awkward, a fifth wheel. Now she considered changing her mind; she wanted to talk to somebody about what Chip had said. Was he just getting revenge for her refusal to go out with him again, or was he really trying to tell her something about the firm, about the partnership decisions? Then she noticed Phil Crisman heading her direction and since there was nobody close enough to hide behind, she escaped to the ladies' room.

Suellen Hancock, an associate a year behind Joan, was standing in front of the mirror repairing her makeup. She looked up as Joan entered. "Hi, Joan. You look like you're on the run."

"Phil Crisman was heading my way for the second time tonight."

"Gee, he's not so bad. At least he doesn't feel you up. Or look at you like he has no idea who you are or how you got admitted to this august company. He does always remember my name."

"What a gift you have for looking at the bright side. I'll try to remember that next time I get stuck between Phil and the shrimp bowl." Joan went to the mirror and grimaced. "I hate how I look in this light."

Suellen nodded and continued with her makeup; then she brushed her hair, washed her hands, and started for the door. "You want to come with us to dinner? It's me and Dan and a few other people."

"No, thanks." She spoke without thinking; it seemed she didn't, after all, want company for dinner. Then it occurred to her that Suellen might know what was going on with Chip, so she added, "Before you go, you haven't heard any rumors about the partnership decisions, have you?"

"Nope. This place is like a sieve on any other issue, but on partnership decisions, forget it. I tried to find out about Bob Smith a couple of years ago when he was up for it, and it was made very clear not only that I couldn't find out, but that I wasn't making friends in asking. So now I don't ask."

Trying to keep the disappointment out of her voice, Joan said, "Okay, thanks anyway. Have a nice dinner." She washed her hands and then looked in the mirror again. Her face looked strange and she thought she might be closer to tears than was safe. "We'd better go home," she told herself.

She took the elevator back down to the thirty-second floor and went to her office to get her briefcase. She filled it even though she knew she wouldn't do any work at home tonight. The briefcase was leather, given to her by Neil on her graduation from law school. Just carrying it made her feel professional and successful. Having it full of work she could do if she wanted to made her feel, in addition, secure. She left the office, turning off her light as she went, and then saw Michael Murphy at the end of the hall, doing the same.

Michael was a partner, about the same age as Chip Lucas. Unlike Chip he was respectful, almost distant, with the women attorneys, businesslike with his secretary. He was a trial lawyer; Joan had worked with him on several cases and had been impressed with how easy he made things look in front of the jury. She had recently discovered that he too was separated from his spouse, and the shared experience had drawn them together several times.

Michael saw her down the hall, smiled, and waited for her to catch up. "You're leaving by yourself? Aren't you going out to dinner?"

"It's all couples. I know nobody else would mind, but I'm not ready for it."

"I know what you mean. The prospect of going out with the crowd my wife and I used to eat with was daunting. You wouldn't

like to go get some dinner with me, would you? I still have to get some work done tonight, but I also need to eat and it would be more pleasant with company. Giovanni's isn't far and it's not so trendy you can't get a seat without reservations."

"That would be nice. I don't think there's a single edible thing in my refrigerator at the moment."

"Great, let's go."

They got their coats and waited for the elevator in silence, Joan wondering if she could ask him about what Chip had said at the party. Or, as Suellen had suggested, would that get her in trouble? She was trying out various questions in her mind when the elevator came. Michael whistled "Joy to the World" as they rode to the lobby, making conversation unnecessary. They walked the few blocks to the restaurant in silence, both of them huddling into their coats against the sharpness of the wind.

Joan still hadn't decided what to do when they got to the restaurant. She let Michael order wine for both of them and studied her menu. When the wine came and they had ordered their dinners, Joan, because she couldn't really think about anything else, told Michael what Chip had said earlier. "I'm not asking you if I'm going to make partner, or if anybody else is—I'm sure you can't tell me that, and I don't want to put you in an awkward position—but I wonder what's going on. Chip isn't the kind of guy who asks how you are for no reason. He doesn't make idle conversation." She wondered, suddenly, if Michael knew about her sexual encounter with Chip. Probably; everybody else had seemed to. She felt the blood go to her cheeks and tried to hide her discomfort by taking a sip of her wine.

Michael looked at her thoughtfully, as if he were trying to decide how much to say. "Things don't look good," he said finally. "I've been involved in some hearings and have missed some of the meetings, so I may not have all the details, but the main thing is, there's a battle going on between the old guard and the young and restless."

"Like Chip?"

"He's one of the most restless. The old guard says the law firm should be a place where there's room for lawyers with different skills and qualities. Translated that means some people don't have to be brilliant because they can bring in business, and others don't have to bring in business because they work hard and don't complain. And a few don't have to do much of anything at all because they're old and have done their share, or just because they're part of the club. The old guard have mostly paid off their mortgages and they get kind of offended if people think they're not being paid enough. Chip and some others, who haven't paid off their mortgages and who bring in a fair amount of business and bill a lot of hours, think they're not getting the money they deserve because of the people who don't bring in business. The question is not just who's going

to make partner this year, but who's going to make partner for the next ten years, and which people who are already partners should be looking elsewhere for jobs. You're a good lawyer—as far as I know, everybody agrees on that—but you don't bring in business and you don't bill extra hours. So, you can see it's a problem."

"It doesn't seem fair. When they hired me they didn't say anything about getting clients. No one has ever said anything in my salary review evaluations." Conscious of how whiny she sounded, Joan stopped. "Oh, never mind. I know a law firm is a business. Where are you in all this?"

"I should be okay either way, since I bring in more than I take out, at least at the moment. The real question would come if Chip loses and then does what he's been hinting at—taking clients and associates and starting a spin-off firm."

"You say it so casually. Doesn't this bother you? It infuriates me. Chip and those guys seem so greedy. I don't think Chip understands anything really about how an organization works. He doesn't understand about cooperation, he doesn't think he'll ever get sick or old and need to depend on somebody else, he thinks secretaries and paralegals don't earn their salaries, and he treats them like slaves if they're not pretty, like concubines if they are (which goes a long way toward explaining why he can't keep a secretary for more than a year). He thinks only money matters. And now it seems if he can't get what he wants he's going to destroy the firm."

"Wow! You're really fired up about this! But I don't think you have the whole story, even if you're mostly right about Chip. He's got support from lots of people, partly because there really is a fair amount of dead wood in this firm—people who not only don't bring in new business but don't do much to keep what we already have. And there's a lot of market pressure on us to keep raising salaries for new associates, which means raising everybody else's salary, too. The money has to come from somewhere."

"Well, maybe we shouldn't keep raising salaries. I certainly make more than enough to live on. I'd be willing to take less if it meant having a better place to work—fewer people like Chip Lucas and more like the old guard."

Michael smiled quzzically. "Would you? How much less? And would it change things if I told you Chip has been one of the people really insisting on hiring lots of women? Or that our major clients are getting dissatisfied with the service they've been getting from the old guard?"

"I'd heard that about Chip and never quite believed it—or thought if it was true, it was probably just so he'd have a wider choice of dates. About how much less I'd take, I guess it would depend at least partly on everybody taking less. I don't know."

"Of course," Michael said, "I shouldn't have asked you that. It's not a fair question. Look, I think our food is here. Can we

drop this? I don't know what's going to happen. If Chip wins and you don't make partner, I'm sure you won't have any trouble finding another job. You'll get good recommendations, certainly from me."

. . .

"Thanks. I know I don't sound very grateful—I am really. I just hope I don't need any recommendations."

Joan spent Christmas Day with her parents, who had not yet reconciled themselves to her separation and had invited Neil to dinner in the hopes of getting the two young people back together. Neil had accepted and when Joan found out, she almost backed out. Because it was Christmas she went anyway, making polite conversation and escaping to the kitchen to do clean-up whenever she could.

It was a relief to get back to work. The week between Christmas and New Year's was always quiet; lots of people were on vacation and those who weren't spent much of their time returning gifts and walking from office to office chatting. New Year's Eve she spent with a friend from law school who talked the whole evening about what a disaster her social life was. Joan drank too much champagne, slept late, and then felt groggy and irritable all New Year's Day. She woke up on January second eager to get to work where everyone would be back from vacation and there would be familiar routines and impersonal problems to deal with. Not until she entered the revolving doors to the building did she remember this was the day she would find out whether the firm would make her a partner. Her eagerness was instantly replaced by dread, a superstitious conviction that if something bad could happen, it most certainly would.

The receptionist gave her two phone messages when she got off the elevator. "Thanks, Irene. Did you have a good New Year's?"

Irene, a well-coiffed, well-dressed woman in her fifties who was friendlier than she looked, shrugged. "All right. I hope the news is good." She nodded at the messages.

"I do, too." Joan read the messages on her way down the hall. One was from a lawyer representing the plaintiff in a case she'd been working on for five years. The other simply said she should meet the partnership committee in the small conference room at ten o'clock. It was more than an hour until then. She dropped her briefcase in her office, greeted her secretary, got coffee, went back to her desk, and started to look through her mail. She had to read everything six times and even then it didn't sink in.

At ten, she walked down the hall to the conference room, realizing as she got there that she was clenching her fists so tightly her nails had left deep imprints in the palms of her hands. She took a deep breath and opened the door. Chip Lucas was sitting at the center of the long table, flanked by two somewhat older partners who worked in different departments of the firm. She knew what the decision was before Chip opened his mouth and hardly listened as he told her, with many expressions of regret, of confidence in her abilities,

and of respect for her work, that she would not be made a partner.

They seemed to go on and on. Richard Goldman, the man on Chip's left, a securities expert, explained at length about the firm's financial situation and the changes that needed to be made to meet the future. George Trent, the third partner, a tax lawyer who specialized in foreign investments, said it had been a hard decision to make. She was a good lawyer, they knew that. "But we just can't afford to have anybody on board who can't pull their oar full strength."

"I'm sure you understand," Chip said. "Now, about the mechanics of it. We've decided it really wouldn't be fair to expect you to leave right away, so we're giving you until March 31. By then you should be able to finish up or turn over any projects you've got going here, and line up another job. In any case, it wouldn't be good for you or the firm if you stayed longer. Obviously, if you want to leave sooner, we won't stand in the way."

"I'm sure you'll have no trouble," George Trent said.

"And there will be a substantial severance payment," added Richard Goldman.

"Thank you," Joan said, out of reflex.

"Thank you, Joan," said Chip, "for your many years of work for Hill, Ledbetter. We're as sorry as you are that, times being what they are, we simply can't see our way clear to making you a partner. If there's anything we can do for you, just let us know." He stood up and held out his hand to her. She shook it automatically and then withdrew it with a jerk. There was nothing to say so she left and walked back to her office, hardly aware of where she was.

She hadn't really believed it would happen, in spite of Chip's comments at the party, in spite of what Michael had told her at dinner. She'd been at Hill, Ledbetter & French since she graduated from law school. The firm wasn't perfect, but she'd been happy. She couldn't imagine working anywhere else. It seemed a worse blow than the breakdown of her marriage. She had spent more of her waking hours at work than she ever had with Neil, and it had been in large part her sense of competence and value at work that had helped her survive the sense of failure in her personal life. Now she had nothing. No husband, no job.

Joan tried for awhile to do some work. She managed to dictate a letter and she started to read a memo prepared for her by one of the younger associates, but it was useless. She could not focus on what she was reading. She kept wondering what she could have done differently. Should she have worked more hours, joined more clubs in the hopes of finding clients? What if she hadn't gone to bed with Chip? What if she'd gone to bed with him more often? What if she'd been married to someone more presentable than Neil? What if she'd done more work for Chip Lucas, Richard Goldman, and George Trent? She sat at her desk for an hour, trying to remember every

mistake she'd made, every opportunity missed. Finally she gave it up. She looked at her calendar. She had a motion scheduled for 1:30 and a client coming to see her at 3:00. She decided to take an early lunch.

On her way out, the receptionist gave her a questioning look. Joan shook her head and her eyes filled with tears. Irene started to say something but the phone rang. She picked it up and mouthed to Joan, "I'm sorry."

Joan couldn't say anything for fear of crying, so she waved as she got on the elevator. She left the building, intending to walk aimlessly till she felt more under control, but she'd forgotten how cold it was. Even in a down coat and fur-lined boots she had to move fast to keep circulation in her toes and fingers. What she really wanted, she realized suddenly, was to shop, and where she wanted to shop was Marshall Field's, the glamour store of her childhood. It was only three blocks from her building to the store and she practically ran.

Once she was inside she felt herself relax. It was so warm, so familiar; nothing really could go wrong here. She wandered up and down the aisles, bought herself a wool scarf she'd been looking at for weeks, and then started up the escalator on a slow ascent to the cafeteria floor, stopping anywhere that looked interesting. By the time she got to the cafeteria, she had added a small copper saucepan and a silk chemise to her purchases. She knew they weren't the kind of things somebody just fired should waste her money on; she had bought them defiantly, thinking it might be her last chance for luxury. For lunch she was sensible; soup and a cup of yogurt. She started to eat, wondering whether the copper saucepan was so extravagant a purchase that she really had to return it, when she felt someone tap her shoulder. "Sally!" Sally Fleck was a friend from law school who worked at another large downtown law firm. She and Joan often met for lunch.

"Hi, Joan, how are you? What are you doing here? You're not waiting for somebody are you? May I join you?"

"No, fine. I'm thinking about my future."

"In the Marshall Field's cafeteria?"

"It seemed a good idea at the time." Joan smiled wryly.

Sally laughed. "Why are you thinking about your future? Are you pregnant?"

"No, I was fired."

"Fired! Oh, no! Why?"

"Oh, it's complicated, office politics, etc. The bottom line is that I don't bring in enough business. Or maybe the real bottom line is that I didn't keep fucking Chip Lucas, pardon my French."

"Really? Maybe you could sue for sexual harassment."

"I doubt it. The fact that I went to bed with him once of my own free will would destroy my case before any jury. Anyway, I

don't think he could have gotten the others to go along with him if there hadn't been other more legitimate reasons. Hill has always been so conservative, you know, that they're still kind of gentlemanly. They wouldn't approve of taking advantage in that way. I've been told they have the best maternity benefits in town, and I think it's because they think mothers should be with their babies."

"Get back to the subject. What kind of legitimate reasons? From what you've said, they've never had any problems with your work."

"There's a fair amount of obfuscation; what it comes down to, as I said, is that I don't bring in business."

"They've never said that matters, have they?" asked Sally. "I didn't think anybody brought in business at Hill. All the clients are ones you've had for fifty years."

"That's sort of true and sort of not and anyway, times are changing, or so they say. There are a few people at my level who do have clients of their own. Gary Pomeroy went to college with a guy who started up a specialty software company and has us do all their legal work, and Robert W. Higginbottom III has family-related tax work, lots of it, and Jim Walters does sports law for his former teammates."

"Have you noticed there's something funny about that list?"

"I know, they're all men," Joan answered. "But is it sex discrimination that I don't happen to know anybody who runs a business? Even if it is, is it Hill's sex discrimination?"

"That's the trouble with being a lawyer. You can see all the arguments on the other side. So, what are you going to do?"

"I don't know. That's why I'm sitting here thinking about my future. And trying to decide if I should return this $100 copper saucepan that I just bought to make myself feel less desperate."

"I wish you could come with us, but we're not hiring. We just merged with Stason, Fry and everything is up in the air."

"I don't think anybody is going to want me if I can't bring clients with me," Joan sighed. "Which I can't."

"Buck up, old girl. Of course they will. We're just in an unusual situation. You've got experience, and Hill will give you good references, won't they? You'll find something."

But she didn't. She called all her friends from law school, read the want ads in legal journals and newspapers, and sent out letters to a hundred firms and companies. At the end of a month only ten firms had expressed any interest in talking to her and many of those cautioned that they weren't sure they'd be able to hire anyone.

The first interview was with a small firm that wanted to expand its litigation department. Joan talked to the head litigation attorney (there were only two). The interview went well until he told her what her salary would be—a third less than what she had been making and not much more than the usual starting salary for recent graduates. She was surprised, and said so.

"We would expect to make an adjustment after six months if everything worked out, a small adjustment. You have to remember we're a small firm and we'd be taking a chance on you. While you've been practicing for several years, you've always done it under supervision. Firms like Hill can afford to do that. We can't. Here you'd be on your own and you must admit, Ms. Bennett, you don't have much experience of managing cases. If we hired you, it would have to be on a trial basis."

Was this legitimate, or was he just taking advantage of her? "Well, thank you for taking the time to talk to me, Mr. Krug. I'm afraid under those circumstances, I couldn't promise to take the job if you offered it to me."

"We'll be making a decision within a few weeks. You'll let us know, of course, if you want to withdraw."

Joan nodded and said goodbye, feeling both angry and disheartened. She didn't want to look for a job. She had hated interviewing for jobs when she was in law school and had been greatly relieved when Hill, Ledbetter & French had offered her a permanent job after she worked there one summer. It wasn't any easier now. She felt even more on the line, more judged on the basis of who she actually was and what she actually knew, so that to be rejected now felt like a judgment on the merits.

The second interview was even worse. The attorney she talked to was both obsequious and familiar, awed by her credentials and pleased to have the opportunity to patronize her. When Joan realized she wasn't going to be introduced to any of the senior partners, she told him she wasn't interested. At the next firm they told her she'd be expected to spend her time managing documents and supervising the paralegals in one large case that had started three years ago and was expected to last another five. She asked what would happen when the case was over. They told her they'd evaluate that when the time came. She took an instant dislike to everyone she met at the next interview, and the sentiment was apparently reciprocated. By the sixth interview she felt that her smile was plastered on a face that had no relation to the person inside. Out of the ten prospects, she ended up getting two job offers, both involving less interesting work at significantly lower salaries than her job at Hill, and she turned them down. She suspected she was being unrealistically critical about the new jobs and unrealistically nostalgic about her present job, but she couldn't help it. Nothing looked remotely as good as what she had at Hill and she couldn't bring herself to settle for less.

At the end of February, still jobless, she met her husband Neil for lunch. He told her he had a girl friend and wanted to start the divorce proceedings so he would be able to remarry. She smiled and said of course. She was, after all, the one who had insisted on the separation. But it was too much. No one seemed to want her, to have any use for her. She couldn't face the thought of more job

hunting and more rejection in her present state and told herself it would be useless anyway. She would give off the scent of failure. No one would even think of hiring her. And it was easy to pretend business was as usual at Hill, Ledbetter. Although she was starting to turn her work over to other people and she wasn't taking on anything new, she still had plenty to do. The people at work were friendly and supportive. If her colleagues heard of a possible job opening, they passed it on. Suellen stopped by several times to ask if she wanted to go to lunch. Even her secretary, ordinarily a rather sour, unfriendly woman, greeted her cheerfully and did Joan's work more promptly than usual. Joan buried her head in the sand, and when her last day came found she had no job to go to.

At the farewell lunch her friends gave for her she was cheerful about it. It had been a long time since she'd had a vacation. With this time off, she could put her life together, clean her apartment, and really devote herself to the job search. But when she left work for the last time that evening, she cried. And after a week, it became clear that her cheerfulness couldn't stand up to reality. Hard as it had been to go on interviews while she was still at Hill, it was next to impossible now that she was unemployed. There never seemed to be a really good reason to get up in the morning. There was no one to greet her, no creature, even a pet, who looked forward to seeing her each day, no important mail demanding her attention. She felt she might as well not exist.

She had saved enough money to live for a few months if she was careful. Still, no matter how frugal she was and how little she ate, the time would come, and soon, when she would have to make some money. She knew that, but didn't know what to do about it. She felt she had tried all her options.

Neil called to find out when she was going to get a lawyer. They needed to come to some agreement about the house and furniture; he wanted to be able to make plans and he couldn't as long as things were unresolved between them. His lawyer was Fred Silver and from now on, she should contact Mr. Silver about anything related to the divorce.

"I don't have any money. Maybe I'll represent myself."

"Get yourself a lawyer, Joan. I'm telling you for your own good. It would probably be better for me if you didn't, but I don't want to take advantage of you; I wouldn't want it on my conscience."

She swallowed several mocking retorts to this expression of concern and said she'd take care of it. She knew of Fred Silver; he had made a name for himself doing celebrity divorces. Much as she hated to admit it, Neil was right. She would have to get a lawyer, but she didn't know where to start. It seemed ridiculous—she'd been practicing law for six years—but it was true. No one at Hill, Ledbetter did divorces. Someone at her friend Sally Fleck's firm might handle divorces, but Joan certainly couldn't afford their fees.

After a lot of humiliating phone calls, all of which seemed to require her to explain to still employed colleagues that she'd lost her job and had no money, she finally got a name, Francine McCallum, a sole practitioner specializing in divorce. Ms. McCallum's office turned out to be on the second floor of a three-story building in a neighborhood in transition between blue-collar middle class and poor. There was a tiny waiting room pretty much filled by two shabby armchairs and a coffee table with a wide assortment of well-thumbed magazines. The inner office was clean and neat but there were no book-lined walls, no thick carpet, no original art on the walls. There were several scratched olive green file cabinets, one of which was topped by a wilting poinsettia, some framed circus posters on one wall, and a diploma from Harvard Law School on another.

Francine McCallum did not look like a Harvard lawyer. In the harsh fluorescent light her face was sallow, her hair dull. She had none of the sleek, polished, well-fed look that Joan thought of as normal among her colleagues. And she spoke with a marked Southern accent. Francine offered her a cup of coffee and when she refused, poured herself a cup from a coffee machine in the corner and dosed it with lots of cream and sugar. She brought it to her desk, sat down, and gestured to Joan to sit in the chair on the other side. Then she pulled out a pad of paper and started asking questions. When she asked where Joan worked, Joan hesitated. Finally she said, "Nowhere at the moment. I used to work at Hill, Ledbetter & French, downtown, but I haven't worked there since March and I haven't found anything else yet."

Francine—she had asked Joan to call her by her first name—looked up from her notepad. "Where've you been looking?"

"Everywhere," Joan said, rather defensively. "Big firms, small firms, corporations. There's nothing."

"Ever thought about trying it on your own?"

"Not for very long. I didn't bring in business there—that's why they got rid of me—why should I think I could bring in enough to feed myself?"

"How about free-lancing? I know a lot of lawyers like me sometimes need research they don't really have time to do. In fact I've got a friend now who has a brief due next month and could use some help. He asked me, but I just don't have the time. Would you be interested?"

"I don't know. I'd have to think about it." The idea of working, even in this part-time, haphazard way, both pleased and frightened her. She'd already gotten used to having nothing to do and was afraid employment would mean being judged and found deficient. She looked down at her lap, twisting her hands in a gesture of helplessness and then, realizing what she was doing, looked up at Francine. "What am I saying? Of course I'm interested, though I'm

not sure why you're doing this. You don't even know if I'm any good or not."

"If you lasted how many—six?—years at Hill you probably know a little something. Besides, I want to get paid, don't I?"

For the next six months, while the two lawyers negotiated a settlement, Joan paid her most pressing bills by doing research projects and briefs for friends of Francine, most of them sole practitioners. In her six years at Hill, one of the largest firms in the city, Joan had seldom been exposed to lawyers like the ones she was working for now and had let her sense of superiority go unchallenged. She was surprised to find that, although some did very well and some merely scraped by, none of them wished they worked for the big firms.

Joan usually spent one or two afternoons a week in Francine's office, using her computer to write up the research Joan had done. Sometimes at the end of the day, if Francine didn't have to stay to see clients, they would go out together for a hamburger or a pizza. At first they talked about business—the divorce or perhaps a legal issue that one or the other wanted to talk through. Before long they discovered they liked each other. In the early fall, when Neil's lawyer finally sent over signed copies of the settlement agreement, it seemed natural to both of them to celebrate by going out to dinner.

"So what are you going to do now?" Francine asked after they had placed their orders.

"The same thing I have been doing, I guess. The divorce won't change my job situation."

"I know, but don't you think you ought to start looking for something more steady?"

"That's funny coming from you," Joan said. "I don't think of your job as particularly steady. You know, I don't really understand how you do it—you and all the other people I've been doing work for. For me, one the best parts of practicing law is having people to talk to about what I'm doing. At Hill, nobody ever worked on a case all by herself. And of course now I talk to the people I'm doing the work for. Doesn't it make you nervous to have to make all the decisions yourself, with nobody even to check them against?"

"Of course it does. It's the hardest part."

"You wouldn't have had to, right? You went to Harvard and I bet you did really well. You could have gone anywhere. And here you are doing blue collar divorce work."

"I'm sure there's more than one reason I'm here and not at a firm like Hill. Part of it, for all of us I think, is liking not having other people tell us what to do. But another reason has to do with what I've seen happen to people who work at the big firms. I worked at a big firm for a year, so I know first hand. People think of lawyers as hired guns, willing to say anything, but it's not true. Lawyers come to really believe in their clients. I mean, I know intellectually there are two sides to every divorce and the person who

files for it isn't necessarily the one who wants it most. I know my clients aren't saints. But I still am on their side. Most of them are women and by now, I just see things their way. If you had stayed long enough at Hill you'd have seen the world your clients' way. The difference is your clients were big corporations. Somehow it's much more scary than the hired gun hypothesis which implies you still have your own values and ideas somewhere inside."

"So you're telling me I was in danger of losing my soul at Hill, I'm lucky I got out when I did, and I should be careful how I make my money because it will change me?"

"God, did I say all that?" Francine smiled. "When my mamma taught me never to get preachy? She hates my being a lawyer. Women are supposed to run things, only not where anybody can see they're doing it. But that's neither here nor there. Let's get back to what you're going to do. Will you try the big firms again?"

Joan laughed. "After what you just said? I'm surprised you even ask. Seriously though, I don't think I will. If somebody from Hill or from Cooley & Watson or another of the big firms called me up tomorrow and offered me a job, I might take it just because I need the money, but I don't think so. Anyway, they're not likely to, so I don't need to worry about it. One or two of the people I've done work for have asked if I'd be interested in more steady work. They'd give me access to their offices—so I could stop using yours—and put me on a retainer or something, and then I'd do their work first. Still part time. I haven't really thought about it seriously. I'm still just living day by day, I guess."

The waitress brought them their salads. They ate in silence for a few minutes. Then Francine said, "You're not going to stop using my office are you? Even if you work part time for somebody else?"

"Not anytime soon. Who else would let me do it for free?" They grinned at each other and went back to their dinners.

A Civil Rights Agenda for the Year 2000: Confessions of an Identity Politician

FRANCES LEE ANSLEY*

Good morning. It is a pleasure and honor to be here today,[1] and most especially a pleasure and honor to share the same platform with Anita Hill. The topic for this symposium would be a daunting one, even if it were not for her and the other illustrious company at this table with me: "Civil Rights and the African-American Community: Setting the Agenda for the Year 2000."

I am not an African-American, and I will not be speaking to you today from the perspective of the African-American community, although I am firmly convinced that my own material and spiritual well-being—and the material and spiritual well-being of my children, and of each person sitting here in this room—is intimately bound up with the well-being of that community. I will be speaking instead from my own perspective and my own situation: that of a European-American, a female, someone of an age to have been imprinted in a profound way by seeing in action the Jim Crow institutions of my Southern childhood, imprinted in another profound way by seeing a generation of black leaders, intellectuals, and ordinary citizens in breathtaking motion around me, a feminist, a teacher of law students, a person who has counted myself a part of legal and social struggles for justice—for men and women of color, for women of all races, and for people (of all races and both genders) whose economic resources consist only of their increasingly uncertain ability to sell their labor to others. I expect I will have less to say to and about

---

* Associate Professor of Law, University of Tennessee College of Law. A.B., 1969, Radcliffe-Harvard College; J.D., 1979, University of Tennessee College of Law; LL.M., 1988, Harvard Law School.

1. The essay that follows is a slightly buffed and expanded version of a talk given at West Virginia University in Morgantown, February 26, 1992, as part of the First Annual Franklin D. Cleckley Civil Rights Symposium. Other speakers at various points during the two-day gathering included James Douglas, Dean of Thurgood Marshall Law School at Texas Southern University in Houston, Dennis Hayes, Assistant General Counsel of the NAACP, Anita Hill, professor of law at the University of Oklahoma, Dr. Benjamin Hooks, Executive Director of the NAACP, and Marilyn Yarbrough, professor of law at the University of Tennessee and visiting professor of law at West Virginia University College of Law, 1991-92. I wish to thank Professor Franklin Cleckley and all the faculty, staff, and students at West Virginia University who worked so hard to make the symposium a memorable event.

the African-American community than to and about the white community.

Finally, I come with few pat answers about what the civil rights agenda for the year 2000 should be. I find myself alternately confused, enraged, inspired, dejected, and sometimes even hopeful about the situation presently confronting us civil rights partisans.[2]

Now I need to pause for a moment and ask you to observe what I just said: "us civil rights partisans." I have noticed that when people use "we" or "us" or "our" or other such terms, it is important to recognize just who it is they mean, who it is they seem to be including and excluding, whether consciously or unconsciously, in their circle. I have concluded that such noticing has value, whether the person is the drafter of a constitution (remember "We the people"?), or a President of the United States (remember "We are not in a recession"?), or some obscure law professor from the University of Tennessee. So just now you (and I!) should notice that when I talk about "us" and "we," I mean "we racial justice advocates: we who believe that the dismantling of the structures and patterns of white supremacy in America has not been achieved and who believe that something needs to be done about that fact."

The definition I just gave may in fact exclude some people in this audience. If so, those of you who are excluded, please accept my warm welcome as witnesses to this part of the conversation. "We partisans" won't be the only "we" or "us" I'll talk about this morning, because I am part of many different groupings, just as each of you is. At this point, however, I intentionally have chosen to talk to "us civil rights advocates." Because I am not here to try to persuade anyone that entrenched racial hierarchy is a prominent feature of our society, or that the vast majority of Americans stand to gain immensely from its abandonment. I want to press on beyond those admittedly important and sometimes contested points to pose questions that, for me, are harder to resolve.

As we near the end of the millennium we seem to find ourselves at what law professor Derrick Bell has called a "crossroads" in civil rights theory and practice.[3] Many civil rights thinkers of many different persuasions have observed that the civil rights crusade faces a crisis.[4] For members of many communities of color there is a crisis

---

2. For some earlier of my attempts at questions and non-pat answers, *see* Fran Ansley, *Race and the Core Curriculum in Legal Education*, 79 CAL. L. REV. 1511 (1991) and Fran Ansley, *Stirring the Ashes: Race, Class and the Future of Civil Rights Scholarship*, 74 CORNELL L. REV. 993 (1989).

3. Professor Bell taught a seminar at Harvard Law School entitled *Civil Rights at the Crossroads*. *See* DERRICK BELL, AND WE ARE NOT SAVED: THE ELUSIVE QUEST FOR RACIAL JUSTICE xii (1987).

4. *See, e.g.*, Alan Freeman *Racism, Rights and the Quest for Equality of Opportunity: A Critical Legal Essay*, 23 HARV. C.R.-C.L. L. REV. 295 (1988);

in living standards, in quality of life, in education, in unemployment, and too often a crisis in survival itself.

For scholars and legal advocates there is also something of an intellectual crisis. The tools and weapons we and our forerunners fashioned earlier in the fight—tools like the Fourteenth Amendment (designed to assure the rights of citizenship to the newly freed slaves), tools like Title VII of the Civil Rights Act of 1964 (designed to assure open participation in the world of work), tools some people gave their *lives* to forge, tools we sincerely hoped (and sometimes even believed) would cause the structures of racial subordination and domination to tumble down—these tools in many cases lie bent and twisted at our feet. They may even be turned against us now to block or hobble ongoing efforts to ease the racial disparities that continue to plague our society. The very fruits of victory in some cases seem to have become the ashes of defeat.

I want to suggest two reasons this morning why the victory of civil rights law reform has sometimes left such a bitter aftertaste. First, in far too many cases and places, the "victory" of even formal equality is still yet to come. I believe it is crucial, especially for "us white people" (please notice that I have shifted to a different "us" at this juncture—stay alert!), to realize just how much old-style, flat-out racial bigotry and unequal treatment is still with us.

Such a realization is something we will have to work at, because in the absence of special luck or special effort, most of us whites simply don't have equal access to adequate information on this score. We can, of course, seek such information out, through reading and study and movie-going and cross-race conversations and through engaging in efforts to change things (which is often the very best way to find out what makes those things tick), all endeavors I highly recommend.

Sometimes, through some association with people of color, we stumble onto information about persistent racist beliefs and disparate treatment. I find myself remembering particular incidents here. One is the racism my brother-in-law found among teachers at the local high school in the district where he and his family live. This racism never came to his attention when his two older boys, who are white, were attending the school. It became all too evident, however, in his dealings with the school when his third, mixed-race child came along.

I think of other incidents too. Now that we are beginning to desegregate the profession of law teaching, for instance, we law

---

Walter E. Williams, *The False Civil Rights Vision*, 21 GA. L. REV. 1119 (1987); Kimberlé Williams Crenshaw, *Race, Reform, and Retrenchment: Transformation and Legitimation in Antidiscrimination Law*, 101 HARV. L. REV. 1331 (1988); Linda S. Greene, *Race in the 21st Century: Equality Through Law?*, 64 TUL. L. REV. 1515 (1990).

professors may now hear stories—by mouth and in print—from African-American and Latino and Asian-American colleagues. Thus, my profession has afforded me the chance to hear, from a fellow teacher in the Northeast, what it is like to commute two hours to work because of her repeated inability to find suitable rental housing. Her white peers have experienced no such problem. The dishonest and depressingly similar conversations she has with landlord after landlord leave their scars.

Sometimes one of my white students relates a similar window of opportunity in his or her own life. One student who had been an undergraduate at Ole Miss recalled inviting an African-American friend down from Nashville for the weekend. He left the friend at home for a couple of hours one afternoon while he went out, returning to find his friend shaken and enraged. Apparently the friend had made the mistake of stepping out of the apartment momentarily for a breath of fresh air. Once outside he was accosted and questioned exhaustively by security guards who simply couldn't believe he might "belong" in that apartment complex. My student's friend was angry but not surprised, whereas the student, a young white man, had learned a brand new lesson.

I had a similar opportunity myself last summer when I made a trip to the United States-Mexico border with a group of women factory workers from Tennessee. We were visiting the border area to see for ourselves what is happening in the industrial zones there where so many United States companies are moving. Our group was mostly Anglo, but one member of the delegation was a black woman, and during part of our trip we traveled with a Latino man who served as our translator. When we stopped at the border the whites in our group watched in amazement as (1) the Latino man was removed from our car, taken off by U.S. border guards to an adjoining building and interrogated alone at length, (2) the black woman was questioned extensively and with evident hostility and distrust about her country of origin, and (3) the rest of us were waved through without a hitch. Had we been traveling without these special "tour guides," my guess is our impression of the border would have been quite different.

We white people thus may have to work at obtaining information and perspectives that others are in a position to observe and verify on a daily basis,[5] but I should not overstate the case. We probably have access to some information on this score that people of color often don't have: we hear the language of other whites who feel they can "speak freely" in our presence. Again, stories from my students

5. Proofreading this text in May, 1992, I feel compelled to mention the special kind of window created by the Rodney King videotape, though I imagine most of my readers are well ahead of me.

have enriched my understanding of this phenomenon. I will share some of them in paraphrase. One told me of the comment he heard in his fraternity: "A black can rush this fraternity; there's not really anything we can do about that, but there will never be a black pledge as long as there is a breath in my body." Another told me of his aunt who is a member of management in a Fortune 500 company, and who supervises a number of black employees. She refers to them as "niggers" when she is home, though she must be more self-conscious about her language on the job. Others tell me stories as well: the child in my Girl Scout troop whose father told her she will not go to a college in a big city, because too many black people would be there. I recall the file clerk at a former job who was surprised I didn't know it was good luck for a white person to rub the head of a young black child.

These and similar attitudes result in hundreds of thousands of human decisions every day, such as decisions not to hire. (Yesterday, Jim Douglas mentioned[6] a recent study where paired teams of identically qualified black and white job seekers tried their luck in two urban job markets. White applicants were three times as likely to get the jobs.[7]) Such attitudes result in other decisions as well: not to rent an apartment, not to grant mortgages in a certain neighborhood, not to promote, not to make eye contact, not to mentor, not to challenge, not to befriend . . . . These hundreds of thousands of decisions help to weave a social fabric, a fabric where the pattern of racial disparity is still being laid down, row by row, day by day, generation by generation.

One reason, then, why the great victory of formal equality with whites has not "worked" very well for African-Americans is that in far too many cases it has never really been tried. Some aspirational goals have been articulated, which is certainly important, and some major inroads have been made. These strong attitudes and decisional patterns persist, however, not as vestiges or remnants or deviant exceptions but as part of the experience of daily life for vast numbers of people. Even formal equality remains an abstract dream in many contexts.

But the frequent failure to achieve formal equality is not enough of an explanation. I think we can and should say more than this about why the victory of civil rights reform has proven so inadequate to the eradication of racial injustice. The problem is more complicated

---

6. Dean Douglas had spoken the previous afternoon, Feb. 25, 1992, on the opening panel of the First Annual Franklin D. Cleckley Symposium, sharing with the audience a collage of his earlier civil rights writings by way of historical and critical analysis of present trends.

7. *See* Julia Lawlor & Jeffrey Potts, *Job Hunt: Blacks Face More Bias*, USA TODAY, May 15, 1991, at 1A.

than simple persistent bigotry and disparate treatment. I believe that both the underlying state of the economy and the underlying state of the law in this country are such that even if true formal equality were achieved tomorrow, the great bulk of African Americans would still be in a perilous condition, and the civil rights movement would still find itself in crisis.

Regarding the economy, I will try to keep my remarks brief. Suffice it to say that the group of the very wealthy is growing, and the group of the poor and near-poor is growing, but what used to be the great group in the middle (typified by the blue collar jobs in industry—mining, steel, auto, furniture, clothing, electronics—which were such an important ladder for black non-professional families into a more secure situation) is shrinking with alarming speed. We are a deindustrializing society whose relative economic strength is in decline. Our government policy has been harnessed to the task of widening the gap between rich and poor, and massive social resources have been shifted from those at the bottom of the social pyramid to those at the top.[8]

As for the state of the underlying law, we have a system of property rights and legal entitlements that results in most people enjoying less security, fewer services, sparser social goods, and fewer cushions against economic disaster than any other advanced industrialized country in the world. The increased pressure for the United States to compete in the global economy will do nothing but exacerbate these tendencies.

If you are black and are not among that top 20% of all Americans who are on the escalator going up, but instead you are part of the bottom 80% that is losing ground, then even true no-faking-it achievement of formal equality would yield only the opportunity for you to stand around with a lot of other people on a glass floor and pray it didn't break. Formal equality would not help you to solve your most basic problems.

What I am saying is two-fold. First, we have won formal equality in principle but continue to learn how hard it is to achieve even formal equality in consistent practice. Second, we are forced to see that formal equality, even if it were honored in practice, would mean little in a society that is in deep economic trouble and that has thus far chosen to guarantee its citizens only the barest of substantive entitlements. Faced with this dilemma, what should "we civil rights advocates" do?

One answer that has emerged from the civil rights movement and from the other struggles for justice that it helped to spawn is what I will call "identity politics." This phrase has been applied to forms

---

8. *See, e.g.*, Sylvia Nasar, *However You Slice the Data the Richest Did Get Richer*, N.Y. Times, May 11, 1992, at C1.

of organizing and forms of political discourse that stress how important it is for subordinated groups of people to mobilize themselves around their own group identity. The recent history of reform movements in the United States has contained a strong dose of identity politics. The civil rights movement itself (especially in its more nationalist manifestations), the women's movement, the gay and lesbian movement, and the movement of the physically challenged are all examples of identity politics at work.

I believe there are some very good and important features of identity politics: the proud identification, study, nurture, and transmission of a group's culture can help to celebrate properly the achievements and sacrifices of subordinated people, to preserve cultural memory, and to create environments that are conducive to human flourishing.[9] Participation in a movement that stresses one's identity and one's bonds with others who share that identity can promote the self-esteem of group members and help them articulate powerfully their concerns and experiences to the larger community. Organizations built around identity politics can create spaces where subordinated people experience a kind of validation, growth, and healthy challenge that may be available to them in no other company and in no other environment.

Further, both history and present observation show only too clearly that certain categories of identity are drastically significant for the distribution of power and resources (and the distribution of powerlessness and pain) in our society. Holding up the lens of race or gender to our world reveals startling patterns that should be noticed and studied. A strategy based on mobilizing members of those groups around visions revealed by those "identity lenses" would seem to have great cogency. For these reasons and more, I view myself as an "identity politician." I am steeped in the habits and outlook of identity politics and believe it offers us important things as we stand here at the crossroads.

In the case of race, racial subordination has been such a lynchpin of our social system for so long and has been built into our lives in so many destructive ways that I believe nothing but a color-conscious movement (and a color-conscious jurisprudence) stands a chance of successfully analyzing or opposing that subordination. That color-conscious movement may find itself entering into much-needed coalitions, but it will and should also find itself insisting that its coalition partners fairly encounter and respond to the tough issues, the history, and the insights afforded by the identity politics of race.[10]

---

9. Margaret Radin introduced me to this evocative phrase. *See* Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1987).

10. This is why the fight to defend affirmative action is so important, for example. It forces an examination of the problems with the ideology of "colorblind-

I believe similar dynamics are at work around the politics of gender. Similar needs exist for independent spaces for feminist work and for an independent women's movement that can prod and challenge its allies to encounter the tough issues, the history, and the insights afforded by the identity politics of gender. I'm sure each of you can think of other examples of groups that could benefit from identity politics.

Dangers are involved, however, in identity politics. Whenever identity is an issue, for example, defining membership in the identity group unavoidably becomes an important task. People involved in identity politics may find themselves spending a lot of intellectual and emotional energy on questions of who is "in" and who is "out." Policing the boundaries can sap people's energy and tax their relationships with others, requiring a kind of "defense budget" for "identity security" that may not be the best use of precious resources.

Related to the problem of military spending in identity politics is the problem of categories. Like all such constructs, the categories of contemporary United States identity politics can distort our vision and the way we think. (The trouble is, of course, while we can't think clearly *with* categories, we can't think *without* them either. This paradox remains unresolved.[11])

One of the main distortions created by categories in this context is that identity politics creates difficulties in coping with people who fall into two categories at once, like people who are both black and female. Of course, a moment's thought will reveal that *all* of us human beings fall into two (and more!) categories at once. Therefore, at least in some ways, identity politics must create difficulty in coping with each and every one of us.

I want to draw your attention to two dynamics in particular that I believe are problematic in the way identity politics handles multiple categories. First, those of us who involve ourselves with identity politics have a tendency to treat the different "identities" a person has as somehow separable from all their other possible identities and also from some generic humanness we all have in common. This is the kind of thinking that leads to questions like, "Which is more important to you, that you are black, or that you are a woman?"

I sometimes think of this as my File Drawer Problem. It has invaded my office space in a very real way. I have one drawer in my filing cabinet labeled "WOMEN" and another labeled "RACE." This makes a certain amount of sense, but I run into all kinds of

---

ness" as it presently functions in public discourse. *See* Neil Gotanda, *A Critique of "Our Constitution is Color-Blind,"* 44 STAN. L. REV. 1 (1991).

11. I am indebted to Angela Harris for putting this difficulty in a particularly elegant light. *See* Angela P. Harris, *Race and Essentialism in Feminist Legal Theory*, 42 STAN. L. REV. 581 (1990).

problems when I file things. Where should I put information relating to the problems of Latina women in South Texas, for instance? Should I create a Latina file for my WOMEN drawer or file the information in the Latino file in my RACE drawer? If I file it in the RACE drawer, am I not implicitly saying that the problems of Latinas at the border are best thought of as racial? Isn't that also inaccurate and misleading, because I know many of their problems are directly tied to their identity as women? If I put the information in the WOMEN drawer, am I not muting that incredibly important part of the problems of Chicana women that springs from their identity as brown people?

Furthermore, if I put the Latinas in my WOMEN drawer, that leaves me with a Latino file in the RACE drawer. What am I supposed to put there? If I put in it anything that relates to Latinos that is not explicitly related to women, am I not giving basic humanity to the men while reserving some special, modified, qualified, different-from-plain-old-Latin status for women? You can see I have a real problem!

My own sense, although it is hard for me to hold onto, is that the "gender part" of a person and the "race part" of the person are not layers that can be thought of as separable strata.[12] A colleague of mine[13] has suggested that perhaps we should stop thinking of things like race and gender as separate layers stacked on each other. She believes it may be more helpful to think of these aspects of identity as "enzymes" that interact with other identity aspects, altering them (and being altered by them) in deeply constitutive ways. They are transformative and interactive rather than add-ons.

While a person's identity may be multi-faceted when seen in this way, that does not mean it is segregable. Asking someone whether her race or her gender is more important would be like asking a molecule of water whether its oxygen or its hydrogen is more important. One familiar with water's elements and properties would recognize this as an incoherent question. After all, if you took away either its oxygen or its hydrogen, you would have no water molecule left. An African-American man is not just a generic male layered over with a stratum of generic blackness. His race has "done something" to his gender identity, and his gender has "done something" to his racial identity. The two form an inseparable whole.

Meanwhile, I have no idea what to do with my file drawers (or with the related problems I find in having to run back and forth

---

12. It is Regina Austin I remember first articulating this for me. *See*, Regina Austin, *Sapphire Bound!*, 1989 WIS. L.J. 539 (attacking "icing on the cake" thinking).

13. Professor Martha Mahoney to be precise, to whom I owe this and other insights.

between two different floors of the library when I want to browse in the holdings on black women). I fantasize that the answer probably lies in a Data Base somewhere in heaven, where all the information I want to collect sits constantly suspended in a sort of humming, never-static matrix, just waiting for me to ask it to reveal itself along one axis or another, but always provisional, partial, perspective-dependent, and contextual. This will require further technology, I suppose. The real problem is what to do with our *minds*, which tend to operate too much like file drawers.

Those of us trying to think about, reason about, and act on these matters of identity still are left with, and need to be aware of, the distortions of categorization that may occur when we try to build a civil rights vision based on identity politics. We do violence to people's multipleness and complexity. We blind ourselves to occurrences that do not fit our categories and that are obscured when we look at a situation with only one lens. We build walls that keep us in our places, even in the process of protesting the injustice of our position.

A second and related problem with categories of identity, a problem at which I hinted when I told you about my File Drawer Problem, is the strong tendency for each category to carry within itself an unstated norm, and for that norm to reinforce and mirror some of the very inequities that the civil rights movement set out to overcome.[14] I am afraid that last sentence may be hard to follow. Perhaps I can best explain by giving an example.

At one point in the development of the feminist movement, white feminists launched a campaign against rape. The idea was to tell the story of sexual violence from a woman's point of view, to redefine the law of rape in a way that was mindful of women's welfare, and so on. This was a terribly important campaign. Those of us in law and legal education are painfully aware of the shameful record of non-enforcement, the need for reform of evidentiary practices, the scandalously high under-reporting rate associated with this violent and traumatizing crime, and related problems. "We law teachers" should thank feminists for bringing these matters into the public consciousness. I want to talk just now, however, about three things "we white feminists" left out of our early accounts of rape:

> —We did little investigation, and spoke very little, of the long and special history of sexual abuse of black women at the hands of white men. This is an important part of the history and dynamics

14. On this point, I especially appreciate the fine work of Kimberlé Crenshaw, who has productively explored these patterns. *See* Kimberlé Williams Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics*, 1989 U. CHI. LEGAL F. 139.

> of rape, and the early white feminist account was impoverished by its relative absence.
>
> —We did little investigation and spoke very little, of the racist use of the rape charge against black men as an instrument of racist terror during long stretches of our national history: a practice in which white women were often complicit, and a setting in which those women could hardly be described as the victims of a prosecutorial process biased against conviction.
>
> —We did little investigation, and spoke very little, of the sexual abuse of black women at the hands of black men, therefore missing entirely an additional burden and constraint often borne by black rape victims seeking security and redress. These victims often experience deep ambivalence about invoking law enforcement authority against a black man because of what they know about racial politics and about the police.

In other words, white women confidently spoke of "We women." We announced that "we" had a whole special set of problems in regard to rape. Upon closer examination, however, the "we" of those initial analyses was not really "we women," it was "we white women." The unstated norm hidden in the term "woman" was in that case "white." Black women's experiences were left out of this account, and the account itself suffered from a narrowness and parochialism that weakened it for everyone. Fortunately, black feminists have been willing to take up this issue and discuss it, and they have provoked an extremely productive reassessment, at least in many quarters of the women's movement.[15]

Another example of an unstated norm occurred in a class I teach on race and gender matters. One day I had asked the class to compare the events, movements, and ideologies that led to passage of the Fifteenth Amendment with those leading to passage of the Nineteenth Amendment. At one point in our far-ranging discussion an African-American male student said something like, "Women have not had to endure the sheer inhumanity that went along with race discrimination and that we blacks have had to bear."

There are a couple of interesting things about this remark. First, you will probably not be surprised that several women in the class wanted to argue against the implication that "women" have not had to bear "sheer inhumanity." Some of the oppressions experienced because of gender are, after all, about as sheer and about as inhumane

---

15. Here I owe a debt to black feminists Angela Davis and Paula Giddings, and to white feminist Elizabeth Spelman, among many others. *See* ANGELA Y. DAVIS, WOMEN, RACE AND CLASS (1981); PAULA GIDDINGS, WHEN AND WHERE I ENTER: THE IMPACT OF BLACK WOMEN ON RACE AND SEX IN AMERICA (1984); ELIZABETH V. SPELMAN, INESSENTIAL WOMAN: PROBLEMS OF EXCLUSION IN FEMINIST THOUGHT (1988).

as one could imagine. Also, this question of rank-ordering oppressions can be a troubling one.

I believe the student's remark had an even more interesting feature, however. Listen to it again: "*Women* have not had to endure the sheer inhumanity that went along with race discrimination and that *we blacks* have had to bear."

The clear implication of that remark is either that all "women" are white (otherwise some of them too would have had to endure race discrimination) or that all blacks are men (otherwise it would be nonsensical to say that "women" as a group can be neatly separated from "we blacks.") The result of this unconscious train of assumptions is the erasure of black women from the mental picture. This way of conceptualizing the problem leads both white women and black men to assume their black sisters away. Black women find a home in neither file drawer.

This puzzle, this utterly unintended slant in my student's language, reveals again the problem of the assumed norm. In the category "women" as used by some feminists in the rape campaign, the assumed norm was white. In the category "blacks" as used by my student, the assumed norm was male. You will note that in each case the assumed norm was that of the dominant half of a polarity: white over black, and male over female. Thus in the very act of claiming one's own identity, one may create another hierarchy that suppresses or erases somebody else, especially if the commitment to one axis of identity is strong enough to block one's ability to see other axes that are also at work.

But that is not all. There is a third problem as well. In addition to the problem of overconcern with boundaries that sometimes accompanies identity politics, and in addition to the difficulty identity politics engenders in accounting for and supporting the multiple, complex identities which people actually do have, the turn to identity politics as a solution to the civil rights crisis can set different groups against each other—groups who ought to be making common cause. It can divide the large group of people whose interests lie in serious change into warring factions resentful and distrustful of each other, worried that any attempt to empathize with the situation of another may threaten the sense of their own identity they have worked so hard to build.

The foregoing discussion reveals some very real problems I perceive with identity politics despite my own position as an "identity politician" of sorts. Now I want to argue that the structure of civil rights law as it has evolved has promoted identity politics in ways that sometimes have been very positive, but at other times have produced real problems in the achievement of meaningful social change.

Our civil rights law is centrally built around the notion of membership in a victim group—what we call a "protected group"

in Title VII doctrine and a "suspect class" in Fourteenth Amendment lingo. In other words, to have a cause of action under much of our civil rights law, one must assert a cognizable "identity," and beyond these special categories of recognized victimhood, one has no grievance.

I first began to appreciate this situation as a limit and a contradiction in the area of employment law. When I had just begun to work for legal services years ago, a worker came in for an interview one day. This worker was convinced by her own astute reading of the signs that she was about to be fired. Her supervisor had taken a personal dislike to her, her children had been through a series of illnesses that had caused her to miss work on several occasions, she had recently been subjected to petty disciplinary actions when it seemed to her that other workers were not similarly disciplined, and her supervisor (a female) had begun criticizing her publicly in ways that seemed both unwarranted and calculated to make her lose her temper. I felt she was right to be concerned about losing her job.

The difficulty, however, was trying to explain to her about The Law. She had this idea that the law would protect her against arbitrary actions by her employer. I asked her if there was a collective bargaining contract at her workplace—if she was represented by union. She replied, "No" (statistically the most probable answer she could have given, of course, here in the U.S.A. in the late twentieth century). I asked whether she had an individual contract of some kind. "No," she laughed (appropriately enough, because she was a blue collar production worker). I asked whether any type of employee manual set out the terms and conditions of her employment. No, she said. I then had to explain that in the absence of a collective or individual contract to the contrary, her employer could fire her for "good cause, bad cause, or no cause at all."

At first she refused to believe me. I explained again. Then she challenged me by reporting that her cousin had once won his job back after he was fired. It seemed her cousin had been the only black man in his department, and his supervisor had repeatedly given him the dirtiest, most dangerous work. After the cousin complained to management about his treatment, the supervisor fired him, but through an anti-discrimination complaint he had been successful in winning back his job. Well, she had me. I had to back up and correct my earlier explanation. Our law was more complex than I had originally told her. Actually an employer could fire her for good cause, bad cause, or no cause at all, except: an employer could not fire her because she was black. There was one particular type of bad cause that the law had put off limits.

My client pressed on. How about the woman that her mother knew who had gotten back pay for having been fired when she refused to sleep with the boss? Well, I hastened to explain, that was another type of unfairness that our law prohibited. An employer

could not fire her because she was a woman, and the courts had developed the view that if someone was fired because she refused to grant sexual favors, that would count as gender discrimination. I went on to explain some other identity groups that might be protected, and some other types of unfair treatment that might be actionable under our anti-discrimination law.

The trouble was that the experiences of this black woman didn't really fit a race discrimination or sex discrimination mold. We could have probed and stretched and perhaps made out a case, but she didn't believe her experiences resulted from race or gender animus. She felt they were individually and personally motivated, arbitrary, and unfair. She felt she should have some recourse and basic job security, and she could not fathom why the law would protect her from one type of arbitrary treatment and not from the others.

Another recent experience reminded me of the strangeness of our anti-discrimination law. A friend of mine served last term on a grand jury in a mid-sized Southern city. One part of that jury's job was to tour the local jails and report on what they found. In this particular locality there were two jails: one city and one county. One of the jails was significantly more comfortable, less depressing, and more spacious than the other. Recently the authorities at the city and county had gotten together and decided to put the women prisoners from both the city and the county together in the less desirable location because that facility was the smaller of the two, and there were fewer women prisoners. The impact on quality of life was, in my friend's view, significant. What was strange for us, however, was to think about how the law might apply to this situation. Many times in the past prisoners had been assigned to this prison through sex-neutral criteria. It seemed likely to my friend and me that a complaint of sex discrimination by the grand jury could succeed in forcing a return to some form of the old days, with a reassignment of prisoners along sex-neutral lines. Exactly the same number of prisoners, however, would still suffer from inadequate quarters. It was hard to feel much victory in such an accomplishment, even though one bad type of arbitrariness would undeniably have been removed from the system.

One important unspoken message of much anti-discrimination law is that our legal, social, and economic system is basically sound and just. This law suggests there is an admitted problem with some social behavior that deviates from this sound and just norm, behavior that makes "arbitrary," "irrational" classifications such as those based on racial or sexual identity. Anti-discrimination law holds out the promise that the legal system can and will eliminate those particular types of arbitrary and irrational deviations. The very promise suggests its corollary, however: that beyond the "fixable" deviations, the law will not and should not intervene.

There were and still are many arbitrary irrational classifications based on race and sex, of course. In my view, however, even if those

were eradicated, plenty of arbitrariness would be left. As Dr. Hooks so wryly reminded us last night, "Most of the folks you know had absolutely nothing to do with how they got here."[16] I wonder, today in West Virginia, how we explain to ourselves who it is that gets born to a college professor, and who to a coal miner? Who is it that is born to a family lucky enough to have a miner at work and who to a single mother struggling to stretch a welfare check? Who is it that is born to a Charleston chemical company executive and who to someone in McDowell County hoping for a job in one of Appalachia's new industries: perhaps burying garbage shipped in from New Jersey or guarding prisoners shipped in from Washington, D.C.? These inequities remain untouched by American anti-discrimination law.

I want to make it clear that I am not blaming the persistence of these non-racial, non-sexual inequities on the selfishness of identity politicians, although those who complain that race and gender issues are the provenance of "special interest groups," or those who oppose affirmative action sometimes seem to suggest as much. Quite the contrary. For example, people of color frequently have litigated to expand the law beyond the suspect-classification, identity-politics branch of equal protection and to strengthen the other, "non-identity" branch of Fourteenth Amendment equality, which is rooted in fundamental rights. Likewise they have lobbied repeatedly for legislation that would benefit more whites than blacks, such as increases in AFDC benefits, food stamps, and the like. Far too often they have enjoyed far too few allies in these endeavors.

If we are searching for causes of "non-racial" problems that beset disadvantaged groups in America, we might well conclude that the stubborn racism of large segments of the white electorate has been the most crucial one, a racism that has prevented those segments from making common cause with people of color. This deep racial divide has been a major reason why the United States lags astoundingly behind other industrialized countries in basic indices of human welfare, such as infant mortality, universal availability of health care, employment security, and adequate education.

So I do not blame us identity politicians for these other kinds of inequities. But I do want to exhort us to action. Identity politicians, and I include myself in that category, must see beyond the lens of their own group identification.

In drawing these remarks to a close, let me suggest one way we might start the kind of process I envision. I am searching for an approach that would not require the discarding or transcending of one's own identity, but rather the deepening of it. Here is my

16. Dr. Benjamin Hooks had given the keynote address for the First Annual Franklin D. Cleckley Symposium the evening before, and had offered his reflections on the past, present, and future of the civil rights movement.

proposal: all of "us identity politicians" who have chosen to identify with a social group and to engage in political activity around that identity, should consciously and as a matter of principle consider the perspectives, the experiences, and the political needs of those members of our identity group who are least privileged. We should conceive of our problems and design our reform strategies with their needs and perspectives firmly in mind.

In other words, a woman like me (a self-identified feminist, white, employed, presently-abled, American, heterosexual, and in a two-parent, two-wage-earner family) needs to investigate how things might look from the points of view of women who are, for instance, black, brown, poor, alien, ill, single, lesbian, third-world, battered, unemployed, or all of the above. Viewing women's problems from those perspectives, I believe, will complicate matters but it will often suggest fruitful answers to strategic questions. Seen from this vantage point, for instance, the goal of helping my sister attorneys crash the glass ceilings at their law firms seems less compelling than universal health insurance, free day care centers, battered women's shelters, and family leave, not to mention development of a responsible industrial policy that aims at sustainable growth both in the United States and for our neighbors in the South. Marilyn Yarbrough was modeling something of this approach for us yesterday, when she urged us to evaluate school choice plans by the standard of how they would affect poor children of color.[17]

In a recent discussion about these matters with a Canadian colleague about this way of approaching identity politics, I was challenged to examine my own assumptions about the meanings of privilege and disadvantage. Some First Nations people,[18] she said, have pointed out to her the ways in which they feel more richly endowed than those in mainstream culture and have urged her to rethink some of her own contrary notions. The point was a provocative one, with which I am still struggling. Nevertheless, there is something I mean here and want to emphasize: certain groups of women have less access to resources, fewer ways to make themselves heard or felt by others, more chances of being marginalized as deviant from a presumed norm, and more likelihood of suffering material deprivations. It is these women I am suggesting "we feminists" should place at the center in our visions and strategies.

You may notice something about the implications of this approach. This conscious and explicit valuing of what some of "us Christians" refer to as "the least of these,"[19] this suggestion that

17. Professor Yarbrough had delivered a talk the previous afternoon on school choice programs and their implications for the African-American community.

18. Some readers may not be familiar with this name for native peoples in the Americas. It is in widespread use in Canada.

19. *Matthew* 25:20.

identity politicians should privilege the least privileged among them, has an interesting side effect: it often suggests coalitions beyond the original identity circle and therefore an expansion beyond the particular identity group in which the project began and in which it remains rooted. What intrigues me is that one arrives at this point by thinking deeply and inclusively about one's "own" group. In the case of feminists, for example, I would argue that points of connection and common cause with men are suggested by thinking deeply and inclusively about the vibrant and elusive category "women".

I will close now by simply leaving it with you. These problems and dilemmas are ours to solve. I invite you to think about your own identity and about the categories of belonging and exclusion that have helped to define you. I invite you to think about those who share some aspects of your identity but not others, and to think especially about those "at the bottom" of whatever category you have chosen for your focus or whatever efforts and institutions in which you find yourself. How might the world look from their perspective? I invite you all to think about each other and about those not here in today's circle at all. Much lies before us.

Students of liberation theology are familiar with the notion of a preferential option for the poor. The story is too long to be explored fully here, but a capsule statement follows.

The "Medellin documents," which emerged from the second plenary meeting of the Latin American Bishops' Conference (CELAM) in Medellin, Columbia, in 1968, comprise a founding statement for liberation theology. *See* PHILLIP BERRYMAN, LIBERATION THEOLOGY 22 (1987) (describing the Medellin documents as "the Magna Carta"). Eleven years later CELAM met again in Puebla, Mexico, and declared the bishops' continuing commitment to the Medellin meeting's "clear and prophetic option expressing preference for, and solidarity with, the poor." *Id.* at 43. They continued, "We affirm the need for a conversion on the part of the whole church to a preferential option for the poor." *Id.* at 43-44. *See also* DEAN WILLIAM FERM, THIRD WORLD LIBERATION THEOLOGIES: AN INTRODUCTORY SURVEY (1986).

In the world of legal scholarship, Mari Matsuda has been one of the most insistent that the fact of multiple perspectives suggests partisanship, that it calls for choice. *See* Mari J. Matsuda, *Looking to the Bottom: Critical Legal Studies and Reparations*, 22 HARV. C.R.-C.L. L. REV. 323 (1987); Mari Matsuda, *When the First Quail Calls: Multiple Consciousness as Jurisprudential Method*, 11 WOMEN'S RTS. L. REP. 7 (1989).

WOMEN, PEACE, AND VIOLENCE: A NEW PERSPECTIVE

STEPHANIE A. LEVIN*

Although we've assembled to discuss "Women's Perspectives on Peace," I want to begin by raising a somewhat heretical question: Do women even *have* any special perspective on peace? My claim is that we have arrived at a critical turning point on this issue.

In the past, women have generally been excluded from any significant role in the making of public policy, so they've had little responsibility for social violence, especially for the violence of war. This outsider status has helped to shape a special perspective on peace. Now, however, while full equality remains a long way off, women are playing a much larger role in public life, including as participants in the making of international and national security policy.

Does this recent entry of women onto the public stage mean that we will move in the direction of a more peaceful world? Unfortunately, I do not think that the answer necessarily is yes.

Even as women come to play a larger role in policy-making, the struggle over the terms of their participation must continue. Will women's greater public role mean assimilation into the violent and unjust practices that now exist, or can those practices themselves be transformed?

In all fairness I feel I should dedicate these remarks to Margaret Thatcher, because it was in thinking about her that I first realized so starkly that being a woman does not necessarily mean being compassionate, especially peace-loving, or humane. Nor is Margaret Thatcher unique. While one can try to explain away women like Thatcher or, to give another example, Jeanne Kirkpatrick, as exceptions to the rule, I think this is a misguided temptation. It may be

---

* Assistant Professor of Law, Western New England College School of Law. B.A., Barnard College, 1967; J.D., Northeastern University, 1978; LL.M., Harvard, 1987.

This piece was originally presented as a talk at the conference "Women's Perspectives on Peace: A Time to Cast Away Stones," held at the University of Tennessee, Knoxville, Tennessee, on March 14, 1992. My appreciation to Barbara Stark for her inspired organizing efforts, to my fellow panelists Nadine Taub, Rosalind Hackett, Barbara Stark, Martha Morgan and James Scott Sessions for their stimulating and enlightening remarks, to Lynn Thomas Strauss and Micki Fox for making everything flow so smoothly, and to Frances Lee Ansley for incisive comments and a home away from home in Knoxville.

true that such women only reflect the type of person most likely to rise to power under existing conditions, or the generally conservative temper of their times. But we have to take seriously the fact that such women exist, and that thousands of others, less celebrated, agree with them. If we don't, I think we are dodging a reality we need to face in talking seriously about women's perspectives on peace.

Jane Addams, the pioneer settlement worker, feminist, and peace activist,[1] focused on this issue early in this century. "I do not believe that women are better than men," she wrote. "We have not wrecked railroads, nor corrupted legislatures, nor done many unholy things that men have done; but then we must remember that we have not had the chance."

Women have not had the chance. We have been prevented from exercising power, we have been excluded from government councils. We have been relegated to the realm of home and family. While men were trained in the arts of war and statesmanship, we were taught the less-valued skills of nurturance and support. These realities helped form our special perspective on peace. But now they are changing.

The invitation to this Conference pointed out that there was a gender gap between those who opposed and those who supported the Persian Gulf War *before* it started.[2] But once it began, this war was notably different from any previous one in the scale of women's participation.

Thirty-five thousand women served in the Persian Gulf, not only as nurses and support personnel, but on the front lines. Two were taken prisoner and eleven were killed, five in action.[3] News media reported favorably on female service members leaving their young children behind as they set off for active duty a continent away.[4] Magazines featured glossy photos of women in combat gear, calling

---

1. For a recent essay arguing Addams has not been given her due as a social thinker and theoretician, see Jean Bethke Elshtain, *A Return to Hull House: Reflections on Jane Addams,* in POWER AND OTHER JOURNEYS 3 (1990). *See also* JANE ADDAMS, THE SECOND TWENTY YEARS AT HULL-HOUSE (1930); JANE ADDAMS, TWENTY YEARS AT HULL-HOUSE (1910); JANE ADDAMS, NEWER IDEALS OF PEACE (1907); JANE ADDAMS, DEMOCRACY AND SOCIAL ETHICS (1902).

2. The invitation to the Conference opens with this paragraph:

Before the attack on Baghdad, there was a chasm between those who supported military action and those who opposed it—a chasm not of age, income, race or geography, but of gender. Why were women so opposed (at least initially) to the bombing? Does peace mean the same thing to women that it does to men?

3. *See* Barbara Kantrowitz, *The Right to Fight*, NEWSWEEK, Aug. 5, 1991, at 22, 23; Jon Nordheimer, *Women's Role in Combat: The War Resumes*, N.Y. TIMES, May 27, 1991, at 28.

4. *See, e.g.*, Linda Siteman Appleton, *Mothers at War*, SPRINGFIELD UNION-NEWS EXTRA, Feb. 9-10, 1991, at 1E.

them "conquering heroines."[5] And NEWSWEEK reported after the war that 52% of Americans felt that women should be assigned to ground-combat units, while only 44% were opposed.[6]

This participation gave renewed impetus to a movement by women both inside and outside the military to dismantle all remaining barriers to women's equal participation in the armed forces.[7] For women, the main barrier to equality in the military is the so-called "combat exclusion."[8] Until recently, this gender distinction was largely accepted as "natural." For example, when the United States Supreme Court decided *Rostker v. Goldberg*,[9] it upheld males-only registration for military service on the grounds that only men could serve in combat. At that time, only a decade ago, women were not prepared to challenge the combat exclusion itself, but now they are.

After a hard fought battle, women have already won the right to be admitted to the service academies,[10] and they have distinguished themselves. They served in Panama and the Persian Gulf. This past summer, Congress repealed the statute that prevented female pilots from flying combat missions.[11] The services are still resisting these changes, but I believe that eventually they will occur.

One of the arguments made by those who seek to end the combat exclusion is that in order to be a full citizen, you must have a right to fight. Fighting for your country, some say, is the ultimate act of patriotism, and until women as well as men have the opportunity to engage in that conduct, they will not be recognized as truly equal citizens.

This same argument was made in connection with African-Americans and the military, and it's instructive to compare their experience in gaining equal rights in the military with that of women.[12] At one

---

5. This characterization comes from A. Craig Copetas, in the photo essay *Conquering Heroines*, MIRABELLA, June 1991, at 26.

6. *Opinion Watch*, NEWSWEEK, Aug. 5, 1991, at 23, 27 (reporting results of NEWSWEEK poll).

7. It should be noted a parallel battle for acceptance by the military is being waged by gay and lesbian servicemembers and their advocates. *See, e.g.*, William Rubenstein, *Challenging the Military's Antilesbian and Antigay Policy*, 1 LAW & SEXUALITY 239 (1991) (reviewing ALLAN BERUBE, COMING OUT UNDER FIRE: THE HISTORY OF GAY MEN AND WOMEN IN WORLD WAR TWO (1990)).

8. For further discussion of the effort to end the combat exclusion and its implications, see Stephanie A. Levin, *Women and Violence: Reflections on Ending the Combat Exclusion*, 26 NEW ENG. L. REV. 1201 (forthcoming 1992).

9. 453 U.S. 57 (1981).

10. Waldie v. Schlesinger, 509 F.2d 508 (D.C. Cir. 1974) was the first case to successfully challenge the exclusion of women from the Air Force and Naval Academies.

11. *See* Annette Fuentes, *Equality, Yes—Militarism, No*, THE NATION, Oct. 28, 1991, at 516.

12. For an interesting exploration of this comparison and related themes, see

time, blacks were entirely excluded from military service, or confined to menial tasks. It was said about them, as is now said about women, that they weren't competent enough or brave enough to fight. Even when they were admitted to the military, they were kept in segregated units, and not until after World War II were they finally integrated on equal terms. Now, of course, Colin Powell, an African-American, is one of our highest-ranking military policy officials.

Surely someday a woman will occupy a position equivalent to Powell's, or to General Schwartzkopf's. In terms of equal rights for women, this may be something to celebrate. But unless the institutions themselves have changed, unless the very meaning of "national security" has changed, females in these positions are unlikely to contribute to a special perspective on peace.

The entire relationship between women and violence is in the process of changing. Until very recently, women were brought up to have a sharply different relationship to violence than men. Men were trained to be able to handle violence; they were expected to be comfortable inflicting it, at least under appropriate circumstances, like on the football field or in the military. Women, on the other hand, were brought up to avoid violence. Under all but the most extraordinary circumstances, they were expected to shy away from engaging in it.

This differing relationship to violence was also reflected in legal rules, in areas ranging from the disparate treatment of boys and girls in school athletics, to women's exclusion from registration, the draft, and combat. But now women have been demanding, and winning, admission to previously all-male domains of violence. They've claimed the right to participate in the sports arenas, in the police forces, in the military academies, and now, in combat itself. As a result, women are moving from being excluded outsiders to active participants in systems of social violence. This raises a central question: Will women now participate equally in the infliction of social violence? Or will there be a demand to transform the very terms on which these institutions operate, moving them in a more nurturing direction?

Fifty years ago, Virginia Woolf asked a similar question in her provocative essay *Three Guineas*.[13] British women had just earned the legal right to study at the universities and enter the professions alongside of men. Woolf celebrated this newly won equality, but believed that it created a dilemma. Behind women lay a patriarchal system which devalued their intelligence and confined them to home

---

Kenneth Karst, *The Pursuit of Manhood and the Desegregation of the Armed Forces*, 38 UCLA L. REV. 499 (1991). On African-Americans, the military, and black citizenship generally, see MARY FRANCES BERRY, MILITARY NECESSITY AND CIVIL RIGHTS POLICY: BLACK CITIZENSHIP AND THE CONSTITUTION, 1861-1868 (1977).

13. VIRGINIA WOOLF, THREE GUINEAS (1938; 1966 reprint).

and family; this was to be rejected. But ahead of them lay a public world racked with pugnacity, militarism, and greed.

Woolf was writing at a time when Europe was again on the verge of war. Women were, as Woolf put it, "between the devil and the deep sea" faced with a "choice of evils."[14] They didn't want to return to the old patriarchal system. But neither did they simply want to join the public world as it was. Was there, Woolf wondered, another answer? "How," she asked, "can we enter the professions and yet remain civilized human beings; human beings, that is, who wish to prevent war?"[15]

Woolf recognized that the answer to this question required a transformation of the terms under which the professions were practiced. This is an extremely difficult undertaking. Once, feminist suffragists believed that granting women the vote could itself solve the problems of militarism and exploitation. By bringing their special perspectives to the ballot box, women would create a more humane, more compassionate political order. Sadly, we know that this promise was not so easily fulfilled.

The contemporary wave of the women's movement presents us with another opportunity to try to transform the terms of public life. If we want a more genuinely peaceful and nourishing society, we must be attentive not to miss our opportunity again. We cannot automatically assume that more women means less violence.

The struggle to transform public institutions is very difficult because they exert such powerful pressures on us to conform to them. For example, when I went to law school after doing community work, my friends and I sought to change the legal profession, to make it less competitive, less pugnacious, less brutal in its working conditions. To some extent, the profession has been forced to adapt to the flood of women into it. But I see too many people feeling forced to conform to oppressive norms of the profession, rather than being able to create change. It is surprisingly hard to speak out and say: "No, I won't play by these rules."

The same thing is true in many workplaces. Can you, to give one small example, support a woman, or a man, who must regularly leave work half an hour early to collect a child? Or is this still so frowned upon that people are afraid to be supportive because their own jobs are on the line? There are real risks involved in speaking out for the position that nurturing values should get equal credit with competitive and aggressive values. You may be labelled "not serious enough," "not committed to your work," "not tough enough to cut the mustard."[16] But if we are to move beyond a violent world,

---

14. *Id.* at 74.
15. *Id.* at 75.
16. For a disturbing look at how these standards are wielded to enforce

we must take these risks. We must go beyond notions of equality that end in the equal right to inflict violence, and insist on a more humane public realm.

I pointed out earlier that fighting for your country has often been viewed as the ultimate test of citizenship, and that exclusion from that arena has been one basis for denying the excluded full citizenship. But why should violence be the primary test of citizenship? Armed defense of one's community and one's country may sometimes be necessary. But it is equally necessary to shelter, to provide food, to give comfort—those tasks which Barbara Stark has argued should be recognized as "nurturing rights."[17] If some would claim that you can't be a citizen unless you fight, why not answer that you can't be a citizen unless you nurture?

One goal of the women's movement has been to question the sex segregation of these activities; to insist that fighting must not be confined to men, nor nurturing to women. But breaking down the gendered nature of these categories must be a two way street. Women's increased assimilation into the violent institutions of the society has to be matched by an equal determination to enlarge the nature of those institutions. Our model of public life has been distorted for too long by the primacy of aggression and conflict. We must not be afraid to challenge this model, and to insist on new conceptions of public citizenship that do not have violence at their core.

---

conformity in the workplace, see MARJORIE HEINS, CUTTING THE MUSTARD: AFFIRMATION ACTION AND THE NATURE OF EXCELLENCE (1987).

17. *See* Barbara Stark, *Nurturing Rights: An Essay on Women, Peace, and International Human Rights*, 13 MICH. J. INT'L L. 144 (1991).

COMMENT

Pregnancy in the Workplace—Sex-Specific Fetal Protection Policies—*UAW v. Johnson Controls, Inc.*—A Victory for Women?*

INTRODUCTION

For more than a century women have fought against policies and legislation designed to protect them and their unborn children from hazards in the workplace because these policies have often been used to exclude women from certain desirable jobs.[1] In recent years, it has become increasingly common for industrial companies to exclude fertile women from working in jobs that could expose them to toxic chemicals or radiation.[2] The companies claimed the purpose of these exclusionary policies was to protect the unborn (and unconceived) children of these women,[3] as well as to protect themselves from

---

* I want to thank Professor Frances Lee Ansley for her suggestions on earlier drafts of this Comment and for opening my mind to a variety of subjects that go beyond the scope of this Comment.

1. *See* Mary E. Becker, *From* Muller v. Oregon *to Fetal Vulnerability Policies*, 53 U. CHI. L. REV. 1219, 1221-22 (1986). Even though many women oppose "protective" legislation because of its exclusionary potential, there are also many women who actively support protective policies that improve women's working conditions. *Id.* at 1222 n.10.

2. Companies that have recently excluded women from jobs involving exposure to toxic substances include: Allied Chemical, American Cyanamid Co., ASARCO, B.F. Goodrich Co., Bunker Hill, Delco-Remy, Dow Chemical Co., Eastman Kodak, Environmental Protection & Aeration Systems, Inc., Firestone, General Motors Corp., Goodyear, Gulf Oil, Monsanto, Johnson Controls, Inc., Olin Corp., St. Joe Minerals Corp., Sun Oil, Union Carbide. *See* Becker, *supra* note 1, at 1226; Marcelo L. Riffaud, Comment, *Fetal Protection and* UAW v. Johnson Controls, Inc.: *Job Openings for Barren Women Only*, 58 FORDHAM L. REV. 843, 843 n.3 (1990); Eva M. Auman, Note, *Excluding Women from the Workplace: Employment Discrimination vs. Protecting Fetal Health*, 55 MO. L. REV. 771 (1990).

3. The goal of protecting the fetus by excluding women from hazardous jobs assumes the adverse health effects from these chemicals or other hazards are transferred by the mother to the fetus in utero, rather than, for example, by paternal exposure and transmission. Problems with this assumption are that reproduction in women has been studied with much greater intensity than in men, creating an illusion

potential tort liability in the event these workplace hazards caused harm to any potential child.[4]

Some commentators have suggested that by imposing sex-specific "fetal protection policies," employers not only disregard the rights and autonomy of women but also effectively bar women from an important group of higher paying male-dominated jobs.[5] Thus, imposition of fetal protection policies can detrimentally affect the economic status and medical benefits of women to the extent a woman might not be able to provide adequately for herself and her family, including prenatal care for her unborn children.[6] Ironically, these policies might selectively protect a few potential children while actually causing harm to unborn and living children, as well as to the women themselves.[7]

The propriety of these sex-specific fetal protection policies was litigated with differing results at the appellate level,[8] and on March 20, 1991, the United States Supreme Court decided *UAW v. Johnson Controls, Inc.*[9] The issue presented to the Supreme Court in *Johnson Controls* was whether an employer can exclude all fertile female

---

women and fetuses are at greater risk; and, studies often fail to control for paternal exposures. Joan E. Bertin, *Reproductive Hazards in the Workplace*, *in* REPRODUCTIVE LAWS FOR THE 1990s 277, 297 (Cohen & Taub ed., 1989). It has been argued,

> workplace . . . hazards are often ignored or minimized except when they involve female aspects of reproduction. Then, employers act quickly to exclude fertile or pregnant women workers on the basis of preliminary, inconclusive, and sometimes speculative information. . . . Once the proof is in, chemicals that are particularly harmful to men . . . have been *banned*, while women have been barred from working around chemicals suspected to cause fetal harm.

Id. at 281 (emphasis added).

4. *See, e.g.*, UAW v. Johnson Controls, Inc., 886 F.2d 871 (7th Cir. 1989). Typical fetal protection policies exclude all women with childbearing capacity. Such policies encompass almost all women for most of their working lives, despite age or plans for conceiving children. Bertin, *supra* note 3, at 279; *see also* Shelly Reed Logan, Comment, *Adapting Fetal Vulnerability Programs to Title VII:* Wright v. Olin, 9 EMPLOYEE REL. L.J. 606, 610 (1984); Pendleten Elizabeth Hamlet, Note, *Fetal Protection Policies: A Statutory Proposal in the Wake of* International Union, UAW v. Johnson Controls, Inc., 75 CORNELL L. REV. 1110, 1110 (1990).

5. *See, e.g.*, Mary E. Becker, *Sterile Women Only Need Apply: Fetal Protection Policies and* Johnson Controls, (1991) (This article was published exclusively in WESTLAW, and no citation was provided. Access via "Law Rev." database, query "Johnson Controls."); Becker, *supra* note 1, at 1219; Bertin, *supra* note 3, at 279; Elise Morse, *Reproductive Hazards, A Labor Feminist Alliance*, 16 LAB. RES. REV. 60 (1988); Kathleen E. Nuccio & Robin Sakima Mama, *Effects of Fetal Protection Policies on Women Workers*, 5 AFFILIA 39, 40 (Fall 1990).

6. *See generally* Note, *Title VII—Equal Employment Opportunity—Seventh Circuit Upholds Employer's Fetal Protection Plan*—UAW v. Johnson Controls, Inc., 103 HARV. L. REV. 977, 982 (1990).

7. *See supra* note 6; *see also* Comment, *supra* note 4, at 610.

8. *See infra* Part II and accompanying notes.

9. 111 S. Ct. 1196 (1991).

employees from certain jobs because of its professed concern for the health of the fetuses the women might conceive.[10] The Supreme Court held Title VII of the Civil Rights Act of 1964,[11] as amended by the Pregnancy Discrimination Act (PDA),[12] prohibits this type of sex-specific fetal protection policy.[13]

Although the Supreme Court's decision is an important victory for working women, there are also negative implications. This paper suggests that the Supreme Court's decision in *Johnson Controls* is a victory for working women only within the framework of the issue as presented to the Court (i.e., the outcome being the lesser of two evils). From a broad perspective of women's rights and worker's rights, *Johnson Controls* is illustrative of the inadequate alternatives facing industrial workers, especially working women, in contemporary society.

Part I of this paper is a general overview of the issues surrounding protective legislation, including a discussion of feminist debates addressing this type of protective legislation for working women. Part II sets the foundation for the Supreme Court's decision in *Johnson Controls* by summarizing the recent court cases concerning fetal protection policies. Part III discusses the Supreme Court's decision in *Johnson Controls*, and finally, Part IV explores the reasons why working women should hesitate to celebrate their "victory" in the outcome of *Johnson Controls*.

I. PROTECTIVE LEGISLATION AND "EQUALITY" IN THE WORKPLACE

The disputes about protective policies and legislation have probably existed ever since women began to work outside the home.[14] The more recent fetal protection policies that exclude women from various jobs are similar to the sex-specific protective labor legislation adopted at the end of the nineteenth and early twentieth centuries.[15]

---

10. UAW v. Johnson Controls, Inc., 111 S. Ct. 1196, 1199 (1991). The policy adopted by Johnson Controls, Inc. excluded all women from jobs involving direct or indirect exposure to lead, "except those [women] whose inability to bear children is medically documented." *Id.* at 1200. Women were also barred from "jobs which could expose them to lead through the exercise of job bidding, bumping, transfer or promotion rights." *Id.*

11. 42 U.S.C. § 2000e (1978); *see infra* note 17.

12. 42 U.S.C. § 2000e(k) (1978); *see infra* note 36.

13. UAW v. Johnson Controls, Inc., 111 S. Ct. 1196, 1209-10 (1991).

14. *See supra* note 1.

15. *See, e.g.*, Muller v. Oregon, 208 U.S. 412 (1908). The Supreme Court upheld a statute that prohibited women from working more than ten hours in one day at certain types of jobs, regardless of whether or not the women were pregnant. Thus, women's rights in the workplace were restricted because the court concluded the physical well being of women is an object of public interest in order to preserve the strength and vigor of the race. *See* Becker, *supra* note 1, at 1219.

This earlier protective legislation was preempted by Title VII of the Civil Rights Act of 1964 because it discriminated against women.[16] The Civil Rights Act focused on "equal opportunity" in the workplace.[17] Consequently, the predominant feminist debate over protective policies for women resurfaced in terms of "equal treatment" (women should be treated the same as men), versus "special treatment" (women need special protection) in the workplace.[18] These debates over special treatment in the form of protective policies for women have divided women's groups. On the one hand, many working women would welcome protective employment policies if the policies did not have the potential to be used to exclude and disadvantage women.[19] On the other hand, many commentators have criticized protective employment policies because of female stereotypes[20] and assumptions about women they believe are generated and perpetuated by these policies.[21] Critics claim that the policies assume women have access to the income of a male wage-earner and that a woman's economic contribution to her family is less important than are her biological and domestic contributions.[22] Additionally, the

---

16. Courts have held Title VII applies to state labor statutes and prohibits discriminatory protective policies in employment. *See generally* Becker, *supra* note 1, at 1225; Note, *supra* note 4, at 1110 n.3.

17. Title VII reads in part:

> It shall be an unlawful employment practice for an employer-
> (1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or
> (2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

42 U.S.C. § 2000e-2(a).

18. For a further explanation of the "equal treatment" versus "special treatment" debate, see Wendy W. Williams, *Equality's Riddle: Pregnancy and the Equal Treatment/Special Treatment Debate*, XIII REV. L. & SOC. CHANGE 325 (1984-85) (advocating equal treatment) and Linda J. Krieger and Patricia N. Cooney, *The Miller-Wohl Controversy: Equal Treatment, Positive Action and the Meaning of Women's Equality*, 13 GOLDEN GATE U.L. REV. 513 (1983). *See also* text accompanying notes 32-35.

19. *See, e.g.*, Krieger, *supra* note 18, at 515.

20. For a discussion of the concept of stereotyping, see Anita Cava, *Taking Judicial Notice of Sexual Stereotyping*, 43 ARK. L. REV. 27 (1990); *see also* Catherine MacKinnon, *Reflections On Sex Equality Under Law*, 100 YALE L. J. 1281, 1293 (1991).

21. *See, e.g*, Wendy W. Williams, *Firing the Woman to Protect the Fetus: The Reconciliation of Fetal Protection with Employment Opportunity Goals under Title VII*, 69 GEO. L.J. 641 (1981); Note, *supra* note 6.

22. *See supra* note 5.

policies fail to acknowledge procreative autonomy by treating as irrelevant a woman's birth control practices, age, sexual preference, or intentions not to parent.[23]

Critics also suggest protective policies undermine society's perception of women as competent decision-makers by giving the employer the power to decide what is best for women and their families.[24] Furthermore, women's individual interests are subordinated to their families' interests and to societal interests, and women's rights are subordinated to fetal rights.[25] Finally, employers usually only exclude women from male-dominated jobs rather than hazardous female-dominated jobs.[26]

Protective laws encourage the segregation of women into certain "safe" occupations. This might be acceptable were it not for the fact this segregation is considered to be a major cause of the wage gap between men and women because "women's work" tends to be undervalued and underpaid.[27] In other words, those jobs historically dominated by women (i.e., secretarial or nursing professions) tend to pay less than jobs dominated by men (i.e., truck drivers or hotel clerks) despite comparable education, training, skill, and time invested.[28] Consequently, these protective laws may play a part in the pattern of the increasing concentration of women among the poor.[29] It is ironic that policies that are supposed to protect women and children from harm may actually contribute to the decline in their standard of living.[30]

If legislation was implemented to protect women and their potential children from hazards in the workplace, while at the same time ensuring that women were not disadvantaged or excluded from opportunities to compete at all economic levels, such action would likely be welcomed by many working women.[31] In fact, many women claim this type of legislation is necessary in order for women to fully

---

23. *See, e.g., Johnson Controls*, 111 S. Ct. at 1200.

24. *See supra* note 5.

25. *See supra* note 21.

26. *See* Becker, *supra* note 1, at 1219 & n.1. Women are not typically excluded from working with toxic substances if their job is in a traditionally female-dominated field such as nursing, textiles, offices, beauticians, and lab technicians. *See* Comment, *supra* note 4.

27. DEBORAH L. RHODE, JUSTICE AND GENDER, 161-65, 173-74 (1989); Nuccio & Mama, *supra* note 5, at 39.

28. RHODE, *supra* note 27, at 165.

29. *See* Nuccio & Mama, *supra* note 5, at 40. For more information on issues surrounding the feminization of poverty, see Gregon Mantsios, *Class in America: Myths and Realities*, *in* THE ECONOMICS OF RACE, CLASS, AND GENDER IN THE UNITED STATES 56-69 (1984); RUTH SIDEL, WOMEN AND CHILDREN LAST: THE PLIGHT OF POOR WOMEN IN AFFLUENT AMERICA (1986).

30. *See* Nuccio & Mama, *supra* note 5, at 40.

31. *See generally* Krieger, *supra* note 18, at 516-17.

participate in the labor market.[32] Because the structure of the typical workplace developed according to the traditional roles of men and women, many women are precluded from full participation in the male-oriented labor market given their reproductive and childraising constraints.[33] As one author has described it,

> [w]omen who bear children are constrained by a society that does not allocate resources to assist combining family needs with work outside the home. In the case of men, the two are traditionally tailored to a complementary fit, provided that a woman is available to perform the traditional role that makes that fit possible. . . . Social custom, pressure, exclusion from well-paying jobs, the structure of the marketplace, and lack of adequate daycare have exploited women's commitment to and caring for children and relegated women to this pursuit which is not even considered an occupation but an expression of the X chromosome. . . . Men, as a group, are not comparably disempowered by their reproductive capacities[;] . . . [t]hey are not generally required by society to spend their lives caring for children to the comparative preclusion of other life pursuits.[34]

Consequently, many women would embrace some type of "protective" legislation designed to alleviate these biological and sociological constraints that disadvantage them in the workplace.

The typical debate over "protective" legislation for women centers on the issues of whether biological and sociological "differences" between men and women should be recognized in the law in order to advance equality.[35] In other words, should women receive "special treatment" in the workplace to accommodate their unique capacity to become pregnant and give birth? If the law reflects certain biological differences between men and women, should it accommodate women for certain sociological differences such as the disproportionate labor of childraising that women are traditionally obligated to perform?[36] Is "equal treatment" the better approach to

---

32. *Id.*; *see generally* Robin West, *Reconstructing Liberty*, 59 TENN. L. REV. 441 (1992).

33. RHODE, *supra* note 27, at 161, 172-75. For a discussion of the disproportionate childrearing duties which are traditionally imposed on women, see ARLIE RUSSEL HOCHSCHILD, THE SECOND SHIFT (1989).

34. MacKinnon, *supra* note 20, at 1312-13.

35. *See supra* note 17. Robin West has argued in *Reconstructing Liberty* that achieving equality between the sexes may be a secondary result to the more immediate goal for women: Achieving liberty from disproportionate childraising obligations and freedom from sexual violence. *See* West, *supra* note 32.

36. Some "special treatment" proponents consider the biological differences between men and women as the only relevant difference the law should recognize. Other proponents would include sociological differences such as childraising burdens and economic disparities as relevant differences. For documentation of the disproportionate childraising duties performed by women, see ARLIE RUSSEL HOCHSCHILD, THE SECOND SHIFT (1989).

ensure that "special treatment" is not used to create barriers for women or to economically exclude women?[37]

Advocates of "equal treatment" fear it would be entirely too easy to exclude and disadvantage women in the guise of "special treatment."[38] They argue that treating women differently from men only reinforces the stereotypes of women as marginal workers.[39] "Equal treatment" advocates strive for gender-neutral laws as a means to eliminate sex discrimination and inequality. They argue pregnancy is just one of the many types of physical conditions that can affect workers and thus, pregnancy should be treated within the same structure of rules that apply to workers with other disabilities or illnesses.[40]

Those who advocate "special treatment" for women argue "equal treatment" means treating women as if they are men when there are obvious biological and sociological differences that disadvantage women in a working world designed for men.[41] They argue "equal treatment" only reinforces the status quo of inequality. Thus, advocates of "special treatment" believe in order to advance equality, laws should reflect the biological and sociological differences between men and women. They argue that by recognizing differences, laws can be formulated to accommodate women for their natural "disability" of pregnancy, as well as for sociological conditions that disadvantage women in the workplace.[42]

Currently, these theories of equality based on gender "similarity" and "difference" support the framework of our employment discrimination laws. Under Title VII of the Civil Rights Act of 1964 as amended by the Pregnancy Discrimination Act (PDA), women affected by pregnancy or related medical conditions are to be *treated the same* as other persons not so affected.[43] This language appears to advocate an "equal treatment" theory for pregnant women in the workplace. Nevertheless, the United States Supreme Court has upheld state legislation aimed at giving pregnant women certain employment guarantees that are not offered to other employees.[44] In other words,

---

37. *See supra* note 17.
38. *See, e.g.*, Williams, *supra* note 18.
39. *Id.* at 330-31.
40. *Id.*
41. *See, e.g.*, Krieger, *supra* note 18.
42. *Id.*
43. The Pregnancy Discrimination Act states in relevant part, "women affected by pregnancy, childbirth, or related medical conditions shall be treated the same for all employment-related purposes, including receipt of benefits under fringe benefit programs, as other persons not so affected but similar in their ability or inability to work." 42 U.S.C. § 2000e(k) (1982).
44. California Fed. Sav. & Loan v. Guerra, 479 U.S. 272 (1987). In *Guerra*, the Supreme Court upheld a California statute that required employers to give

the Supreme Court has interpreted Title VII and the PDA as securing a minimum standard of "equal treatment," but individual states are not prohibited from enacting "special treatment" legislation requiring that more protection be provided to pregnant employees than is provided to employees with other disabilities.[45] Under Title VII and the PDA, however, employers and states need not implement "special treatment" policies and legislation, nor must they work toward eradicating the existing disadvantages that confront women in the workplace.

The fetal protection policy adopted by Johnson Controls, Inc. was reminiscent of the older protective legislation that was used to exclude and disadvantage women.[46] The Supreme Court's decision in that case, on a superficial level, was merely a reiteration of Title VII's minimum guarantees to "equal treatment" in the workplace.[47] This guarantee to equal treatment is invaluable; however, it only secured the rights of these women to the limited choice between economic livelihood or healthy children; and one's economic livelihood often determines whether one has healthy children.

Johnson Controls is a victory for women in light of the alternative; however, even this victory is more narrow than is first apparent. In addition, *Johnson Controls* highlights the typical, traditional incompatibility of work and family life that disproportionately disadvantages women in the labor market. Moreover, the outcome in *Johnson Controls* displays a major weakness in our current employment discrimination laws that tend to maintain the status quo of inequality by failing to alleviate the structural barriers that work to limit the alternatives of women in the workplace.

II. Decisions of the United States Courts of Appeal

The Supreme Court's decision in *Johnson Controls* resolved a split in the Circuit Courts of Appeal about the proper application of Title VII and its defenses to the issue of fetal protection policies.[48] Specifically, the split in the Circuit Courts concerned the issue of whether sex-specific fetal protection policies should be evaluated under a "disparate treatment" or "disparate impact" framework.[49] If a policy constitutes "disparate treatment," then employers must show a "bona fide occupational qualification" (BFOQ) defense to

---

workers, disabled by pregnancy, up to four months unpaid leave and a right to reinstatement. Workers disabled for other reasons were not entitled to the same protection under this statute. The Court concluded Congress had not intended the PDA to prohibit preferential treatment, but rather, "Congress intended the PDA to be 'a floor beneath which pregnancy disability benefits may not drop—not a ceiling above which they may not rise.'" *Id.* at 691.

45. *Id.*

46. *See, e.g.*, Muller v. Oregon, 208 U.S. 412 (1908); Becker, *supra* note 1.

47. *See supra* note 17.

48. *See infra* notes 40-46.

49. There are two categories of discriminatory practices that violate Title

justify the discriminatory policy.[50] A "business necessity" defense may be used in "disparate impact" cases and is more easily established than is a BFOQ defense.[51] As described below, conflicting opinions evolved among some of the Circuit Courts of Appeal over the appropriate Title VII analysis that should be used to evaluate these exclusionary policies.

In 1982, the Fourth Circuit Court of Appeals applied a disparate impact/business necessity analysis to uphold a sex-specific fetal protection policy.[52] Similarly, in 1984, the Eleventh Circuit also upheld a fetal protection policy under a somewhat modified disparate impact analysis.[53] The *Johnson Controls* case was before the Seventh Circuit in 1989, which also applied a modified version of a disparate impact/business necessity analysis in order to uphold the fetal protection policy.[54] In 1990, while *Johnson Controls* was pending before the

---

VII. Employment policies that classify workers on the basis of race or sex or other forbidden categories are facially discriminatory and fall within the category of "disparate treatment" cases. "Disparate impact" refers to policies which are facially neutral but have a discriminatory effect on employees. The disparate impact analysis was created by the Supreme Court in Griggs v. Duke Power, Co., 401 U.S. 424 (1971). In *Griggs*, the Supreme Court also created the defense of "business necessity" to coincide with the disparate impact analysis. *See infra* note 51.

50. The BFOQ defense is explicitly provided for in Title VII:

> (e) Notwithstanding any other provisions of this subchapter, (1) it shall not be an unlawful employment practice for an employer to hire and employ employees . . . on the basis of his religion, sex, or national origin in those certain instances where religion, sex, or national origin is a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise . . . .

42 U.S.C. § 2000(e)-2(e) (1982). This BFOQ defense is rather difficult to establish because courts have traditionally interpreted a BFOQ defense very narrowly. *See, e.g.*, UAW v. Johnson Controls, Inc., 111 S. Ct. 1196, 1204 (1991).

51. The defense of business necessity is a judicial doctrine created by the Supreme Court in Griggs v. Duke Power Co., 401 U.S. 424 (1971). Under *Griggs*, once the plaintiff demonstrates the challenged employment practice has a disparate impact on a group protected by Title VII, a business necessity defense may be established if an employer shows the job requirement at issue has "a manifest relationship to the employment in question." *Griggs*, 401 U.S. at 432. In 1989 the Supreme Court made it more difficult for plaintiffs to bring disparate impact cases by requiring the plaintiff to bear the burden of proof throughout the proceeding. Wards Cove Packing Co. v. Antonio, 490 U.S. 642 (1989). The Civil Rights Act of 1991 has since overruled *Wards Cove*. The Civil Rights Act of 1991, Pub. L. No. 102-66, 105 Stat. 1071.

52. Wright v. Olin Corp., 697 F.2d 1172 (4th Cir. 1982). This was the first fetal protection policy case to confront a federal appellate court. Even though the policy specifically barred all fertile women from certain jobs, the Fourth Circuit held the fertility policy was facially neutral but constituted a case of disparate impact justifiable by the defense of business necessity. *Id.*

53. Hayes v. Shelby Memorial Hosp., 726 F.2d 1543 (11th Cir. 1984). In this case the plaintiff was an X-ray technician who was fired from her job when she became pregnant. Instead of applying a disparate treatment analysis, the court created a modified version of the disparate impact analysis which was to be applied

Supreme Court, the Sixth Circuit remanded a case involving an exclusionary policy, and ordered the district court to apply a disparate treatment/BFOQ analysis.[55]

In *Johnson Controls*, the exclusionary policy that confronted the Seventh Circuit barred all fertile women[56] from jobs involving exposure to lead, including jobs that did not involve lead exposure but upon promotion or other transfer could expose the employee to lead.[57] Petitioners affected by this exclusionary policy filed a class action in the District Court for the Eastern District of Wisconsin alleging discrimination under Title VII of the Civil Rights Act of 1964,[58] as amended by the Pregnancy Discrimination Act (PDA).[59] The district court granted summary judgment for Johnson Controls,[60] and the Seventh Circuit sitting en banc affirmed the decision, seven to four.[61]

The Seventh Circuit approved a three-part business necessity defense[62] that the district court had adopted from the Fourth and

---

to fetal protection policies. The Eleventh Circuit held a fetal protection policy "violates Title VII unless the employer shows 1) a substantial risk of harm exists; 2) the risk is borne only by members of one sex; and 3) the employee fails to show there are acceptable alternative policies that would have a lesser impact on the affected sex." *Id.* at 1554.

54. UAW v. Johnson Controls, Inc., 886 F.2d 871 (7th Cir. 1989); *see infra* text accompanying notes 47-62.

55. Grant v. General Motors Corp., 908 F.2d 1303, 1311 (6th Cir. 1990). In *Grant*, the Sixth Circuit adopted the disparate treatment/BFOQ defense as the appropriate standard for resolving fetal protection policy cases. The court reversed a lower court decision that applied a disparate impact/business necessity analysis and remanded so the trial court could apply the disparate treatment/BFOQ test as explained by Judge Cudahy of the Eleventh Circuit in his dissenting opinion in *Johnson Controls*.

In addition to the *Grant* case, another fetal protection policy case was decided while *Johnson Controls* was pending before the Supreme Court. This case, also against Johnson Controls, Inc., was brought in the state courts of California. Based on a California state law similar to Title VII, Johnson Controls' fetal protection policy was struck down. Because the policy applied only to women, the court found the policy was discriminatory on its face and the company failed to establish a BFOQ. The court also noted OSHA standards on lead contained the same warnings for fertile men and women and did not recommend any exclusionary policy. Johnson Controls, Inc. v. California Fair Employment & Hous. Comm'n, 267 Cal. Rptr. 158 (Ct. App. 1990).

56. The policy applied to women with child-bearing capacity, which was defined as "all women except those whose inability to bear children is medically documented." *Johnson Controls, Inc.*, 886 F.2d at 876.

57. *Id.*

58. 42 U.S.C. § 2000e-2 to 2000e-17 (1982); *see supra* note 17.

59. The Pregnancy Discrimination Act of 1978, 42 U.S.C. § 2000e(k) (1982). The PDA amended Title VII to clarify that discrimination based on pregnancy, childbirth, or related medical conditions violated Title VII. *See supra* note 36.

60. 680 F. Supp. 309, 318 (1988).

61. UAW v. Johnson Controls, Inc., 886 F.2d 871, 901 (7th Cir. 1989).

62. Under the Seventh Circuit's business necessity test, a fetal protection

Eleventh Circuit Courts.[63] The Seventh Circuit added an additional requirement to the three-part business necessity test by imposing the burden of persuasion on the plaintiffs for all three steps.[64] Even though the Seventh Circuit agreed with the district court that Johnson Controls was entitled to summary judgment under the defense of business necessity, the court went on to evaluate the exclusionary policy under a BFOQ defense and determined (in dicta) that the policy was also valid under that defense.[65] Relying on the Supreme Court case of *Dothard v. Rawlinson*,[66] which had established a BFOQ defense for particular safety concerns related to the "essence of the business,"[67] the Seventh Circuit reasoned that safety for the unborn is part of the essence of the business of making batteries at Johnson Controls, Inc., and consequently, is a legitimate BFOQ defense to its exclusionary policy.[68]

The four dissenting judges in the Seventh Circuit's opinion in *Johnson Controls* argued the proper evaluation for fetal protection policies is to be made under a disparate treatment/BFOQ analysis.[69] Two of the dissenting judges, Cudahy and Posner, would have remanded for a trial to determine if fetal safety was relevant to the BFOQ exception.[70] The remaining two dissenting judges, Easterbrook and Flaum, determined the company's concern for the safety of the unborn was irrelevant to the operation of the battery-making business, and thus could not fit within the BFOQ framework.[71]

In short, between 1982 and 1989, three United States Circuit Courts of Appeal had approved these fetal protection policies, each

---

policy excluding only women will survive a Title VII challenge if 1) there is a substantial health risk to the fetus; 2) transmission of the hazard to the fetus occurs only through women; and 3) the plaintiff fails to show a less discriminatory alternative equally capable of preventing the health hazard to the fetus. *Id.* at 885; *see supra* note 42.

63. *See supra* notes 43-44 and accompanying text.

64. 886 F.2d at 887-93. The Seventh Circuit cited Wards Cove Packing Co. v. Antonio, 490 U.S. 642 (1989), as authority for the additional requirement. *Wards Cove* was applicable to a disparate impact analysis and required the burden of persuasion to remain with the plaintiff. *Wards Cove*, however, has since been overruled by the Civil Rights Act of 1991. *See supra* note 51.

65. 886 F.2d at 893-94.

66. 433 U.S. 321 (1977).

67. The Supreme Court permitted a bona fide occupational qualification defense excluding women from working as security guards in the contact areas of a high security male penitentiary in which a percentage of inmates were sex offenders. The Court concluded a woman's sex could create a risk of sexual assaults and thus undermine prison security because prison security was the essence or "central mission" of the business. *Dothard*, 433 U.S. at 335-37; *see infra* notes 83-87 and accompanying text.

68. *Johnson Controls*, 886 F.2d at 898.

69. 886 F.2d at 901-21.

70. *Id.* at 901.

71. *Id.* at 908.

using a different analysis, but all three in agreement that policies barring only fertile women were not facially discriminatory.[72] The Supreme Court granted certiorari even before the Sixth Circuit decided, in contrast to the other circuits, that these policies were indeed facially discriminatory.[73] By the time the Supreme Court granted certiorari there was a good deal of attention focused on *Johnson Controls*. Some estimates claimed that the jobs of more than twenty million women were at stake in the outcome of the Supreme Court's decision in *Johnson Controls*.[74]

III. THE SUPREME COURT DECIDES *UAW v. JOHNSON CONTROLS, INC.*

On March 20, 1991, the Supreme Court in *UAW v. Johnson Controls, Inc.*[75] interpreted Title VII and the Pregnancy Discrimination Act in favor of the plaintiffs by forbidding the sex-specific fetal protection policy adopted by the company.[76] The Court held an exclusionary policy aimed at all women capable of bearing children, rather than at fertile men and women, is a facial classification based on gender and potential for pregnancy.[77] This classification based on sex that does not apply equally to the reproductive capacity of the male employees is disparate treatment.[78] Consequently, the discriminatory policy is explicitly prohibited by Title VII as amended by the Pregnancy Discrimination Act (PDA) unless the company can establish a BFOQ exception.[79] Because explicit facial discrimination is disparate treatment, the Appellate Court erred by applying the business necessity defense.[80] Moreover, the Supreme Court held the company's professed concern for fetal health is not sufficient to establish a BFOQ exception to discrimination.[81]

A cursory look at *Johnson Controls* may reveal what appears to be a unanimous decision prohibiting sex-specific fetal protection policies. In fact, the judgment was unanimous only because all of the Justices agreed the Seventh Circuit should not have granted summary judgment, and that sex-specific fetal protection policies must be evaluated under a disparate treatment/BFOQ analysis because they are facially discriminatory.[82] The Justices were not in agreement as to the fundamental issue of whether fetal protection policies could be justified under the Title VII framework. Particu-

72. *See supra* notes 43-46 and accompanying text.
73. 110 S. Ct. 1522 (1990); *see supra* note 46 and accompanying text.
74. UAW v. Johnson Controls, Inc., 886 F.2d 871, 901 n.43 (7th Cir. 1989).
75. 111 S. Ct. 1196 (1991).
76. *Id.* at 1209-10.
77. *Id.* at 1202.
78. *See id.* at 1203-04.
79. 111 S. Ct. at 1204; *see supra* note 41 and accompanying text.
80. *See id.* at 1203.
81. *Id.* at 1207.
82. *See id.* at 1204, 1210. *Compare* majority opinion *with* Justice White's

larly, the majority and concurring Justices disagreed as to the scope of a BFOQ defense. The five-member majority[83] emphasized the restrictive scope of a BFOQ, claiming that the Court had always interpreted the defense narrowly.[84] Unlike the majority, the concurring Justices[85] interpreted a BFOQ defense to allow sex-specific fetal protection policies in certain circumstances where the employer is concerned about fetal health.[86]

In order to determine the scope of the BFOQ defense, the majority and concurring Justices discussed the proper interpretation of "occupational qualification," pursuant to Title VII.[87] Under Title VII, an employer may discriminate on the basis of sex where sex is "a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise."[88] In an attempt to ascertain legislative intent, the majority concluded "by modifying 'qualification' with 'occupational,' Congress narrowed the term to mean qualifications that affect an employee's *ability* to do the job."[89] In contrast, the concurrence interpreted the phrase broadly to mean "*related* to a job."[90] As the majority pointed out:

> According to the concurrence, any discriminatory requirement imposed by an employer is 'job-related' simply because the employer has chosen to make the requirement a condition of employment. In effect, the concurrence argues that sterility may be an occupational qualification for women because Johnson Controls has chosen to require it. This reading of 'occupational' renders the word mere surplusage.[91]

The majority was concerned that the concurring Justices would allow an employer's subjective "idiosyncratic requirements" to suffice as a BFOQ.[92]

The Supreme Court previously recognized sex as a BFOQ in *Dothard v. Rawlinson*,[93] when the safety of third parties was at risk. The Court allowed the employer to hire only male guards in contact areas of a maximum security male penitentiary in which some of the

---

concurrence and Justice Scalia's concurrence. All of the Justices noted the impropriety of summary judgment and that a BFOQ analysis was proper.

83. Justice Blackmun wrote the majority opinion in which Justices Marshall, Stevens, O'Connor, and Souter joined.

84. *Johnson Controls*, 111 S. Ct. at 1204.

85. Justice White concurred in part and concurred in the judgment, in which Chief Justice Rehnquist and Justice Kennedy joined. Justice Scalia filed a separate opinion concurring in the judgment.

86. *Johnson Controls*, 111 S. Ct. at 1213 (White, J., concurring).

87. *See supra* note 41 and accompanying text.

88. 42 U.S.C. § 2000e-2(e)(1) (1982); *see supra* note 41.

89. *Johnson Controls*, 111 S. Ct. at 1204-05.

90. *Id.*

91. *Id.* at 1205.

92. *Id.*

93. 433 U.S. 321 (1977).

inmates were sex offenders.[94] The concern in *Dothard* was that a woman security guard would jeopardize the safety of the other guards as well as the inmates themselves if violence broke out; and also because the woman's sex might create a risk of sexual assaults.[95] Sex discrimination was tolerated in *Dothard* not because of the potential danger to the woman herself, but because the "essence of the business" was to maintain prison security and the fact that the employee was a woman affected her ability to perform her job.[96]

Based on *Dothard*, Johnson Controls, Inc. argued sex discrimination is permissible because fetal health and safety is at risk when women perform certain battery-making jobs.[97] The majority distinguished *Johnson Controls* from *Dothard* because in *Dothard* the Court found a high correlation between sex and the woman's ability to perform her job as a guard in a maximum security prison; whereas in *Johnson Controls*, the majority decided sex had nothing to do with a woman's ability to make batteries.[98]

A second example used by the Supreme Court to illustrate its interpretation of the "essence of the business" test was *Western Air Lines, Inc. v. Criswell*.[99] Even though *Criswell* is an age discrimination case, it is relevant because the BFOQ provisions of Title VII are the same as in the Age Discrimination in Employment Act (ADEA).[100] In *Criswell*, airline employees sued under the ADEA, and the Supreme Court recognized age as a BFOQ when considering the overriding interest of public safety.[101] The majority in *Johnson Controls* emphasized the fact that in *Criswell* "safety concerns were not independent of the individual's ability to perform the assigned tasks, but rather involved the possibility that, because of age-connected debility, a flight engineer . . . might thereby cause a safety emergency."[102]

---

94. *Id.*

95. *Id.* at 335-36. The Court's assertion a woman is not as capable of doing her job as a man because her sex may create a risk of sexual assault has been widely criticized. As described by one author,

> [The Court assumed] that convicted rapists are more likely to rape a female prison guard than to attempt violence against a male prison guard, and that there is no way to ensure prison security short of excluding women from the job entirely. Even so, it is not clear why women should be excluded from this position when there is a risk of being raped outside of the prison context.

Alison E. Grossman, *Striking Down Fetal Protection Policies: A Feminist Victory?*, 77 VA. L. REV. 1607, 1631 n.114 (1991).

96. *Dothard*, 433 U.S. at 335-36.

97. *See generally Johnson Controls*, 111 S. Ct. at 1205.

98. *Id.* at 1205-07.

99. 472 U.S. 400 (1985).

100. *Johnson Controls*, 111 S. Ct. at 1204.

101. 472 U.S. 400, 405-06 (1985).

102. *Johnson Controls*, 111 S. Ct. at 1205.

Again, the Supreme Court distinguished the safety concerns in *Criswell* from those in *Johnson Controls* because the "essence of the business" of an airline is the safe transportation of its passengers that could be jeopardized by age-related debility.[103] Whereas, the "essence of the business" at Johnson Controls, Inc. is battery manufacturing, which is not affected by an employee's sex or fertility, with both fertile and nonfertile women equally able to perform their jobs.[104] As a result, the majority held the professed concern for fetal safety by Johnson Controls, Inc. was not sufficient to establish a BFOQ.[105]

The concurring Justices concluded, "protecting fetal safety while carrying out the duties of battery manufacturing is as much a legitimate concern as is safety to third parties in guarding prisons (*Dothard*) or flying airplanes (*Criswell*)."[106] Moreover, the concurring Justices claimed, based on *Dothard* and *Criswell*, avoiding safety risks to third parties is "inherently part of both an employee's ability to perform a job, and an employer's 'normal operation' of its business."[107] In other words, safety to third parties is part of the "essence" of any business so that a BFOQ is justified whenever safety to third parties is at risk. Thus, the concurring Justices concluded the BFOQ defense was broad enough to permit companies to exclude women from jobs in order to protect the potential fetuses of these women.[108]

Despite the majority's reassurance that the BFOQ defense is narrow, the defense, in practice, has proven to be quite malleable.[109] There are several instances in which the BFOQ defense has been expanded to justify sex discrimination in cases where either gender was *able* to perform the job.[110] Accordingly, it is still possible that the Supreme Court, in the future, might approve sex-specific fetal protection policies as justified within the BFOQ framework, especially given the recent changes on the Court.[111] Likewise, it would not be surprising to see businesses attempt to draft exclusionary policies that

---

103. *Id.* at 1205-06.
104. *Id.* at 1207.
105. *Id.*
106. *Johnson Controls*, 111 S. Ct. at 1213.
107. *Id.*
108. *Id.*
109. For a discussion of the flexible nature of the BFOQ defense and examples of its various applications, see Grossman, *supra* note 95.
110. Sex discrimination has been permitted in situations where privacy interests are implicated, such as excluding male obstetric nurses. *See also id.*
111. Since the decision in *Johnson Controls*, Justice Marshall has retired and Justice Stevens is in poor health; both of these Justices sided with the majority in *Johnson Controls*. It is also important to note the substantial number of lower court judges that appeared eager to uphold fetal protection policies.

are gender neutral but that may have a disparate impact on women.[112] These policies could be upheld under the more expansive "business necessity" defense.[113]

Not only did three of the concurring Justices take the position that a sex-specific fetal protection policy could be justified based on an employer's concern for the welfare of a woman's potential fetus,[114] but all four of the concurring Justices agreed (in dicta) that an exclusionary policy could be justified based on extra costs associated with employing women.[115] Although the five-member majority stated "the incremental cost of hiring women cannot justify discriminating against them,"[116] even they left the question open as to whether a company could justify an exclusionary policy if the extra costs of employing women were "crippling" or "so prohibitive as to threaten the survival of the employer's business."[117]

Although Johnson Controls did not argue it was faced with increased costs from potential tort liability, the majority Justices suggested (in dicta) that in most cases, employers need not be concerned over the costs associated with potential tort liability because Title VII may preempt state tort remedies.[118] In other words,

---

112. It would be possible for employers to implement facially neutral or "benign" policies aimed at both men and women but that have a disproportionate impact on women. For example, a policy could exclude all employees who are able to transmit toxins to their unborn children, and then medical studies could show only women transmit these specific toxins. *See generally* note 3.

113. *See supra* note 51.

114. The Chief Justice, Justice White, and Justice Kennedy supported this position.

115. 111 S. Ct. at 1211-13. In their concurring opinion, Justices Rehnquist, White, and Kennedy agree, "[p]rior decisions construing the BFOQ defense confirm that the defense is broad enough to include considerations of cost and safety of the sort that could form the basis for an employer's adoption of a fetal protection policy." *Id.* at 1212. These Justices assert, "costs are relevant in determining whether a discriminatory policy is reasonably necessary for the normal operation of a business . . . [and] [t]he BFOQ statute . . . reflects Congress' unwillingness to require employers to change the very nature of their operations." *Id.* at 1213. Justice Scalia agrees in his separate concurring opinion that increased costs alone can support a BFOQ defense. *Id.* at 1216.

116. 111 S. Ct. at 1209; *see* City of Los Angeles Dep't of Water & Power v. Manhart, 435 U.S. 702 (1978) (The majority cites this case for the proposition, "extra cost of employing members of one sex, however, does not provide an affirmative Title VII defense for a discriminatory refusal to hire members of that gender.").

117. *Johnson Controls*, 111 S. Ct. at 1209.

118. 111 S. Ct. at 1208-09. In his separate concurring opinion, Justice Scalia suggested:

> any action required by Title VII cannot give rise to liability under state tort law. That assumption, however, does not answer the question whether an action *is* required by Title VII . . . [since] it is perfectly reasonable to believe that Title VII has *accommodated* state tort law through the BFOQ

if state tort law furthers discrimination by preventing employers from hiring women, then state laws will impede the goals of Title VII and may be preempted.[119] In the majority's opinion, however, if the employer has not acted negligently, and has fully informed the women of any risks, the potential for liability is "remote at best."[120]

Three of the concurring Justices disagreed with the assertion that Title VII will likely preempt state tort remedies.[121] These Justices contended that not only is preemption questionable, but it is also difficult to determine what will constitute negligence. Because there is the potential for strict liability, along with the fact that parents cannot waive a cause of action for their child, it would be unwise for employers to dismiss the possibility of tort liability.[122] Accordingly, these three Justices, along with Justice Scalia,[123] might permit exclusionary policies based on the increased costs associated with employing fertile women, including those costs associated with potential tort liability.[124] The Court did not decide either the propriety of a cost-based exclusionary policy or the preemption question, because Johnson Controls did not argue that it was faced with increased costs from tort liability or otherwise.[125]

Because of the recent departure of Justice Marshall from the Supreme Court, it would not be surprising to see a cost-based exclusionary policy reach the Court in the near future. If Justice Thomas sides with the four concurring Justices, then extra costs associated with employing women could justify their exclusion from certain jobs. Moreover, all of the Justices indicated that at some point costs might justify excluding women.[126] No doubt costs associated with cleaning up a hazardous manufacturing process could "threaten the survival of a company." Even a single successful toxic tort action against a company has the potential to elicit "prohibitive"

---

exception.

Id. at 1216 (emphasis in original). Justice Scalia minimized the preemption issue by suggesting that a substantial risk of tort liability is "enough to defeat a tort-based assertion of the BFOQ exception." *Id.*

119. 111 S. Ct. at 1209. The majority stated "[w]hen it is impossible for an employer to comply with both state and federal requirements, this court has ruled that federal law pre-empts that of the States." *Id.* But, if an employer complies with the mandates of state tort law by maintaining a healthy environment for employees (not just male employees) then there does not seem to be any conflict with Title VII.

120. *Id.* at 1208.

121. 111 S. Ct. at 1211.

122. *Id.*

123. *See supra* note 102.

124. *See supra* note 99.

125. 111 S. Ct. at 1209.

126. *See supra* notes 99-101 and accompanying text.

costs for that company. Consequently, it is likely some employers will continue to exclude women from hazardous jobs by claiming it is "too expensive" to include them.

IV. Is *Johnson Controls* A Victory for Working Women?

When considering the alternative of permitting employers to exclude women from certain desirable jobs, *Johnson Controls* is a victory for women because women have more control over their employment status than they would if the fetal protection policy were upheld. On the other hand, the decision in *Johnson Controls* merely assures women of their right to risk good health, including the health of their potential children. From this perspective, *Johnson Controls* represents the lack of adequate alternatives facing workers; working women in particular. This section explores some of the negative implications of *Johnson Controls*, including: The continuing possibility of exclusionary policies; federal preemption of state tort remedies; the Court's characterization of business concerns as entirely distinct from family concerns; and finally, the inadequacy of Title VII to protect the status of women in the workplace.

The limited victory of *Johnson Controls* may be short-lived because there are a variety of ways companies may continue to exclude women from certain jobs. In particular, the Supreme Court left the doors open for employers to implement cost-related exclusionary policies.[127] Other employers may continue to exclude women through facially neutral exclusionary policies,[128] or possibly through narrowly drafted sex-specific policies.[129] Finally, there is always the possibility a changed Supreme Court will opt for a more expansive interpretation of the BFOQ defense to permit exclusionary policies like the one adopted by Johnson Controls, Inc.[130]

If the Supreme Court permits employers to exclude women from certain jobs because of, for example, the increased costs associated with employing fertile women, then women are being penalized on the basis of their gender. These women may be deprived of the means to pay for the health and welfare of their families. In one breath, the concurring Justices spoke of the essence of a business as including safety to employees and third parties,[131] while in the next breath these Justices indicated businesses need not be concerned about paying for the increased costs of providing safe and healthy

---

127. *See supra* notes 99-101 and accompanying text.

128. *See supra* note 112.

129. The policy adopted by Johnson Controls, Inc. was extremely broad. Perhaps, if a policy were drafted to exclude only pregnant women, it might be upheld.

130. *See supra* note 111 and accompanying text.

131. 111 S. Ct. at 1210.

environments to women and their potential children.[132] Likewise, the concurring Justices claim that protecting fetal safety is a legitimate concern,[133] yet employers need not be concerned about the health of the fetuses and children whose mothers have been excluded from employment.

Health insurance is already a luxury by some standards. If women are excluded from the jobs with better pay and benefits, then it would not be surprising to see a rise in infant mortality and health problems in children.[134] Although employers may be able to save costs by excluding fertile women from employment, they also may end up sharing in the increased costs of health care to society associated with, for example, paying for postnatal care of babies whose mothers could not afford adequate prenatal care.[135] Society benefits from healthy children, and employers share in this benefit. Thus, employers should be required to pay the costs of maintaining non-hazardous environments for all workers rather than being permitted to exclude women as if they are only marginal workers.[136]

Although the Justices discussed the possibility that Title VII might preempt state tort remedies,[137] none of them discussed the desirability of this option or its possible ramifications. Accordingly, it is important to question the desirability of foreclosing a company's liability to an injured party rather than leaving the potential for liability intact in order to motivate employers to maintain healthy working environments for both men and women.[138] If employers are not required to maintain non-hazardous workplaces for both men and

---

132. *See* 111 S. Ct. at 1212.

133. 111 S. Ct. at 1213.

134. *See generally* Karen S. Taylor, *National Health Law Program: The Cost Effectiveness of Prenatal Care*, 19 CLEARINGHOUSE REV. 259 (1985).

135. *Id.*

136. It may entail substantial costs to provide a non-hazardous environment for all women. To avoid these expenses, many businesses might prefer to pick up operations and move to countries with fewer regulations and cheaper labor. No doubt, this concern partially fuels the argument for a cost-based exclusionary policy. Because it is fundamental that a business needs to be competitive, an effort must be made, not only to meet the needs of businesses without excluding women, but also to protect off-shore workers as well. Otherwise, this movement toward locations with the fewest regulations and the cheapest labor becomes a degenerating competition of who can withstand the most abuse and poison. The subject of off-shore workers is, of course, huge and complex. Nevertheless, it is relevant and worth mentioning.

137. *See supra* text accompanying notes 102-09.

138. Although companies claim exclusionary policies are necessary to avoid liability for injury to future children of female workers, this is a negligible source of potential liability for these companies compared with all of the hazards involved, such as in the production process, the use of these products by consumers, disposal of byproducts, and various other manufacturing risks. Bertin, *supra* note 3, at 294 & n.51.

women, then the Supreme Court's decision in *Johnson Controls* only guarantees women the right to be treated as equally badly as men.[139] Because studies have shown employers tend to disregard evidence that hazardous work environments affect the male reproductive system as well as the female reproductive system,[140] preemption of state tort remedies, as well as exclusionary policies, ignores the rights of both men and women to work in a healthy environment.[141] Rather than eliminating the company's liability to an injured party, or barring fertile women, it may be more appropriate to focus on the duty of employers to provide a healthy environment for all workers,[142] or to provide non-hazardous alternatives to the susceptible class of employees without penalizing them on the basis of their reproductive functions or gender.

Federal preemption entails some ramifications. For example, if women are required to assume the risks of working in hazardous environments, should mothers be held accountable to the state or their children for any damage that results to their children? If a woman chooses the lower paying non-hazardous job but must give up the economic and medical benefits necessary for adequate prenatal care, should she again be held accountable to the state and her children for any injuries that result? The criminalization of prenatal neglect is now a reality in some states and threatens to encourage an adversarial relationship between mothers and their potential children.[143] The potential to harm the relationship between a mother and her fetus is likely to increase when the mother is penalized for becoming pregnant either by losing her employment status and benefits or by continuing to work in a hazardous position and thus risking injury to her potential child or assuming the risk of a lawsuit brought by her child.

---

139. The Occupational Safety and Health Act (OSHA) requires employers to protect employees from serious harm caused by recognized hazards in the workplace. 29 U.S.C. § 654(a)(1) (1982). In fact, the Supreme Court stated in *Johnson Controls*, "Title VII plainly forbids illegal discrimination as a method of diverting attention from an employer's obligation to police the workplace." 111 S. Ct. at 1209.

140. Bertin, *supra* note 3, at 279-82; *see supra* note 3; *see also Johnson Controls*, 111 S. Ct. 1200 (1991); Comment, *supra* note 4, at 608-09.

141. Bertin argues in *Reproductive Hazards in the Workplace*, *supra* note 3, chemicals will be banned once there is conclusive evidence of harm in male workers, but women are quick to be excluded on inconclusive evidence of possible hazard to a fetus. She suggests this pattern of behavior reinforces the idea employers must accommodate the needs of men but not women, and that men are wrongly presumed to be invulnerable to the effects of chemical exposures until undeniable evidence of hazard is amassed. Bertin, *supra* note 3, at 282.

142. *See supra* note 138.

143. *See generally* Dawn Johnsen, *From Driving to Drugs: Governmental Regulation of Pregnant Women's Lives After Webster*, 138 U. PA. L. REV. 179 (1989); Note, *Maternal Rights and Fetal Wrongs: The Case Against the Criminalization of "Fetal Abuse"*, 101 HARV. L. REV. 994 (1988).

The Court's preemption argument condones a policy of assumption of the risk for fertile women in employment. A further extension of this policy, however, is to require all employees to assume the risks of employment, provided employers inform the workers of the associated hazards that come with making a living. Nevertheless, passage of the Occupational Safety and Health Act (OSHA) requiring that employers maintain certain standards of safety to protect workers from hazardous environments signifies the inappropriateness of an assumption of the risk policy.[144] Certainly, the requirements of OSHA and other workplace regulations should protect fertile women as well as other groups.[145] Compromise solutions exist somewhere between total immunity and unlimited liability for employers. For example, Congress could limit tort awards or extend worker's compensation to cover prenatal injuries.

After *Johnson Controls*, women have the option to choose between an economic livelihood and reproductive health. The Supreme Court illustrated this limited choice when it said "[i]t is no more appropriate for the courts than it is for individual employers to decide whether a woman's reproductive role is more important to herself and her family than her economic role."[146] While the Court made an important and valid statement, at the same time it implied that a woman must make an either/or choice, that work and family are not compatible, and that a woman must choose which is more important to her. Interestingly, just four pages before this statement, the Court expressly recognized that women should not have to choose between work and family; in the majority's words, "women as capable of doing their jobs as their male counterparts may not be forced to choose between having a child and having a job."[147]

Because most men are seldom forced to make this choice, forcing women to choose between their reproductive functions and an economic livelihood is a type of sex discrimination.[148] This type of discrimination arises, at least partially, from the structure of a labor market adapted to the traditional needs of men.[149] Even today, the typical workplace structure is not designed to accommodate repro-

---

144. *See supra* note 117; *see also*, Note, *supra* note 2.

145. At least one court has held the "general duty clause" of OSHA does not bar exclusionary policies, while one other court has held even if fertile women are hypersusceptible, they are entitled to full protection of any comprehensive standards set by OSHA. Bertin, *supra* note 3, at 283-84.

146. *Johnson Controls*, 111 S. Ct. at 1210.

147. 111 S. Ct. at 1206.

148. *See, e.g.*, *Guerra*, 479 U.S. at 286 n.19 (remarks of Rep. Tsongas, the "[PDA] would put an end to an unrealistic and unfair system that forces women to choose between family and career—clearly a function of sex bias in the law.").

149. *See supra* text accompanying note 34; *see also* RHODE, *supra* note 27.

ductive and child-raising constraints that traditionally disadvantage women in the workplace.[150]

An underlying assumption that encourages this type of male-oriented workplace structure is the idea that the business world and family life are somehow unrelated. For example, throughout the majority opinion in *Johnson Controls*, there is little consideration that men, women, and children (or potential children and future generations), are a part of the business world that the business community owes certain responsibilities. For example, the majority states, "[d]ecisions about the welfare of future children must be left to the parents . . . rather than to the employers who hire those parents."[151] Certainly, this statement is valid, but it must be recognized that a parent's decisions are often inextricably woven with employers' decisions —such as an employer's decisions about daycare, health insurance, working schedules, parental leave, or maintenance of a hazardous environment.

The Court's presumption that the business world is unrelated to family life is evident when it begins its analysis by stating, "the bias in Johnson Controls' policy is obvious [because] [f]ertile men, but not fertile women, are given a choice as to whether they wish to risk their reproductive health for a particular job."[152] But why should men or women be forced to risk their reproductive health for a particular job? The framework and substance of the Court's analysis implies businesses need not be concerned with nor accommodate family life.

By advocating fetal protection, the concurring Justices seem to recognize businesses must consider families, including the health and safety of workers and potential children. For example, the concurring Justices state, "common sense tells us that it is part of the normal operation of business concerns to avoid causing injury to third parties . . . ."[153] Nevertheless, this concern expressed by the concurring Justices is superficial because they are unconcerned with the likelihood women and children as a group will be harmed if these exclusionary policies are maintained. Businesses should be concerned with fetal health and worker safety; however, excluding women from desirable jobs without considering their alternatives is not compatible with the responsibilities a business should owe to its workers and to society.

---

150. *Id.* at 173; *see generally* Kathryn Abrams, *Gender Discrimination and the Transformation of Workplace Norms*, 42 VAND. L. REV. 1183 (1989).

151. 111 S. Ct. at 1207.

152. 111 S. Ct. at 1202. It should be noted one of the representative plaintiffs in *Johnson Controls* was a man who was denied a leave of absence so he could lower his lead level because he had intentions to become a father. Thus, men should also be entitled to non-hazardous alternatives to accommodate their fertility needs.

153. 111 S. Ct. at 1210.

As long as the business world and family life are presumed to be separate entities, then it is unlikely the structure of our labor market will become more accommodating to the needs of families; particularly women and children. And, unless the structure of the labor market accommodates the needs of women, women will continue to be discriminated against by being pressured into limiting their participation in either economic production or reproduction. Although Title VII and the PDA have been instrumental at prohibiting certain forms of sex discrimination in the workplace by requiring that women be *treated the same* as men, this focus on "equality" tends to maintain the status quo of sex inequality because it fails to account for the underlying structural and biological barriers that disadvantage women in our labor market.

Our employment discrimination laws and the debate over equal treatment versus special treatment for working women focus on gender differences.[154] It has become increasingly obvious that theories of equality based on sameness and difference are inadequate, partly because of the problems inherent in defining "sameness" and "difference."[155] Under the "difference" theories, the term "equality" is relative because its definition depends on the objects being compared.[156] In the context of our employment discrimination laws, the term "equality" is given meaning based on the male prototype. Accordingly, when men are not accommodated for their reproductive capacities, then, under Title VII, women need not be accommodated for their reproductive capacities either.[157] The problem with this analysis is evident in the case of pregnancy where there are no pregnant men to which women can be compared and treated the same.[158]

One response to the weaknesses of the "difference" theories of equality is to focus on something besides equality with its difficult puzzle of comparing relevant gender differences.[159] For example, it

---

154. *See supra* notes and accompanying text.

155. *See, e.g.*, MacKinnon, *supra* note 20; Nancy E. Dowd, *Work and Family: The Gender Paradox and the Limitations of Discrimination Analysis in Restructuring the Workplace*, 24 HARV. C.R.-C.L. L. REV. 79 (1989).

156. For example, women are entitled to the same treatment as men, only to the extent that they are like men. Catharine MacKinnon points out that this model of equality where likes are treated alike and unalikes are treated unalike is the same concept of equality that a Nazi author used to justify hierarchy under the Third Reich. MacKinnon, *supra* note 20, at 1287 n.31.

157. Although Title VII only requires this "minimum standard" of equality, the Supreme Court upheld state legislation under Title VII that provided more protection to pregnant employees than to employees disabled for other reasons. California Fed. Sav. & Loan v. Guerra, 479 U.S. 272 (1987); *see supra* note 37.

158. *Id.* The Supreme Court recognized this problem in *Guerra*.

159. For example, one author argues the focus should be on the fact sex inequality exists and thus, she argues any policy or practice that furthers the

may be more appropriate to focus on the "well-being," "welfare," or "flourishing" of women. Changing the focus from equality to "well-being" is especially necessary when dealing with issues such as sexual violence and pregnancy, which are problems that particularly affect women.

Perhaps, the focus should be on women's "liberty" rather than gender equality as Robin West suggests in her essay, *Reconstructing Liberty*, at the outset of this symposium issue.[160] Professor West argues that Congress and the states may have an affirmative constitutional duty to advance women's liberty by protecting women against private infringements of their right to fully participate in society, including the paid labor market.[161] For example, under Professor West's theory, the fetal protection policy adopted by Johnson Controls, Inc. would be subject to constitutional challenge because it is a private infringement of women's rights to full participation in the labor market. She suggests sex equalization should improve as a secondary result of the immediate goal of freeing women from constraints that disproportionately disadvantage women such as sexual violence and child-rearing.[162]

In addition to protecting women from private infringements on their liberties, Professor West suggests that some type of "affirmative protection" legislation would be constitutionally required. In the context of employment discrimination, Title VII is an example of affirmative protective legislation. Title VII, with its focus on equality, protects a woman's "negative liberty." In other words, Title VII, as interpreted by the Supreme Court in *Johnson Controls*, protects a woman's right to be left alone to make her own decision about whether she wants to choose a hazardous job. Another example of affirmative protection legislation would be the Occupational Safety and Health Act designed to protect workers from hazardous working environments. If the OSH Act required employers to protect the health of fertile women and was effectively enforced, then this would be a step towards protecting women's "positive liberty;" or the right to work in a healthy environment. Consistent with Professor West's theory, legislation such as Title VII and the OSH Act could contain provisions to affirmatively protect the "liberty" or the "well-being" of women. These provisions would require the state and private sector to accommodate working mothers and fathers in conceiving, birthing, and rearing the next generation.

---

subordination of women should be presumed invalid. *See generally* MacKinnon, *supra* note 20. *See also* Grossman, *supra* note 95 (application of Catharine MacKinnon's theory to fetal protection policies).

160. West, *supra* note 32.

161. *Id.* at 454.

162. *Id.* at 454-55.

CONCLUSION

Certainly, women do not desire to risk harm to their potential offspring, and thus, for many women, the victory of *Johnson Controls* is rather limited. For most women, employment is an economic necessity; however, biological and sociological barriers often force women to work in more flexible types of jobs that are typically female-dominated, low-wage, and low-status. When the risk of toxic harm to a woman's potential child is added to the other biological and sociological barriers confronting women in the workplace, the fact that women lack the same options as men to fully participate in our contemporary labor market is highlighted.

Although the Supreme Court relied on Title VII and the PDA to prohibit a discriminatory exclusionary policy adopted by Johnson Controls, Inc., the outcome of *Johnson Controls* also exposed the limitations in our sex discrimination laws. These laws are not designed to affirmatively promote equality in the workplace; instead, they tend to maintain the status quo of inequality. The United States Supreme Court's interpretation of Title VII in *Johnson Controls* secures "equal opportunities" for women only if they can "act like men." Thus, women are left with a choice between economic health and reproductive health. Even this limited choice, however, looks good compared to the possibility that the Supreme Court will further limit women's options by allowing employers to exclude women altogether, justified by various interpretations of the Title VII defenses to discrimination. Hopefully, Title VII can function as a safety net or minimum standard of equality to prohibit the use of protective policies that disadvantage or exclude women.[163]

The initial solution to the problem posed by *Johnson Controls* would be to require employers to maintain non-hazardous environments for all workers, rather than being permitted to exclude women. Certainly, the minimum standards of the Occupational Safety and Health Act should apply to men and women alike. If Johnson Controls, Inc., is truly concerned about the welfare of potential fetuses, as it has professed,[164] then the company should accommodate the fertility needs of men and women by offering non-hazardous alternatives that do not penalize them for exercising their reproductive capacities. Moreover, where businesses fail to accommodate reproductive needs, Congress should require this type of "affirmative protection." Otherwise, women truly do not have an "equal opportunity" in the workplace.

JENNIFER MORTON

---

163. The California legislature passed "special treatment" legislation that was upheld by the United States Supreme Court in California Fed. Sav. & Loan v. Guerra, 479 U.S. 272 (1987). *See supra* note 37.

164. 111 S. Ct. at 1207.

Address at Rio Earth Summit*

SENATOR AL GORE

The United Nations Conference on Environment and Development, or the Earth Summit, held recently in Rio de Janeiro, moved the world toward a better understanding of how future progress is inextricably linked to improvement of the environment. It also laid the groundwork for meaningful changes in policies of every nation on the face of the earth to stop the destruction of the global ecological system. For these reasons, I believe the Earth Summit in its largest sense was a magnificent success. Every nation in the world is now thinking about the same challenge at the same time in a new way.

For many years, heads of state, parliamentarians, legislators, scientists, journalists, and many others concerned about the environment have talked with each other about ecological tragedies in one nation or another. And local dialogue has been similar to the old parable of the blind men and the elephant. One of the blind men had the elephant's trunk in his hand. Another had the elephant's tail. And yet another had one of the elephant's enormous legs. And they described to each other completely different animals. Only after enough time had passed and enough communication had taken place was there a realization that they each had a separate part of the same beast.

At the Earth Summit, there was the same realization that the disappearance of the Aral Sea, the burning of the rainforest, the disappearance of living species at a rate one thousand times greater than the natural extinction rate, the tragedy of the Love Canal, the garbage crisis, the oil spill in Alaska, the dead dolphins in the Gulf of Mexico, the dead seals in the North Sea, and the dead starfish in the White Sea are all different parts of the same global crisis.

This crisis was addressed by heads of state in Rio and will continue to be discussed around the world. The challenge is in making the difficult decisions on how to implement this new way of thinking—this new shared recognition that human beings must heal the relationship between our civilization and the ecological system of the earth.

---

* Senator Al Gore, Address at the Earth Summit, Rio de Janeiro, Brazil (June 5, 1992). Taken from a speech to the Global Forum of Spiritual and Parliamentary Leaders on Human Survival. Senator Gore chaired the United States Senate delegation to the Earth Summit.

A number of dramatic changes made it possible for the Earth Summit to take place. For instance, people all over the world now feel that they are part of a single global civilization. We are a community of separate nations, and we shall remain so; but we face the same problems, and we must construct a common agenda for solving those problems. The Earth Summit was the first of many conversations that will take place on a global basis as we endeavor to create that global agenda.

Some remarkable agreements were arrived at during the Earth Summit. Most of the world now agrees that freedom is a prerequisite for solving the global environmental crisis. As people who care about the future of our world, we must make a compact with one another to struggle against the enemies of freedom. Dictatorship is an enemy of freedom. Communism is an enemy of freedom. Ignorance is an enemy of freedom. Corruption is an enemy of freedom.

Racism and sexism, exploitation and oppression are enemies of freedom, and everywhere we look in the world today—wherever the human spirit is crushed, wherever individuals feel powerless and live out their lives in fear that they have no meaning or purpose—human beings and the environment suffer. From Eastern Europe and the states of the former Soviet Union, to Ethiopia, to Tibet, to Haiti, and to South Central Los Angeles, we have an obligation to hear what people are saying and to feel the suffering when we hear the numbers of children who are dying in the world every single day. We have an obligation to link the democracy movement and the environmental movement. We can do so and we must do so.

Many of us have come to believe as human beings that the ecological crisis is fundamentally a spiritual crisis. It is not an accident that in those countries where the environment has been the most devastated, human suffering is usually the worst. Where the human spirit has been pushed down and where hope has been abandoned, we see the erosion of the soil, the cutting down of the forest, and the poisoning of the water and the air.

It is a spiritual crisis because the crisis springs out of the relationship between human beings and the ecological systems of the earth. There have been three dramatic causes of this transformed relationship between human beings and the earth.

The first is the population explosion. After 10,000 generations of human beings, there were little more than two billion people on this earth in 1945. In the last five decades, population has increased from a little over two billion to five and a half billion and in the next five decades, it will rise to at least nine or ten billion.

We must listen when those who best understand the population problem tell us that it is linked not merely to the availability of safe and effective birth control on a voluntary basis, but it is also profoundly linked to poverty, injustice, human suffering, and child mortality. As Julius Nyerere said thirty years ago, the most powerful contraceptive in the world is the confidence by parents that their

children will survive. When we save the lives of children, we make it possible for mothers and fathers to choose to have smaller families. These same experts tell us that the population problem is inextricably linked to the level of education and literacy in the world, especially among women who must be empowered to participate in the decision making.

Let us listen to spiritual and religious leaders when they say aid for developing countries must not be conditioned upon policies related to birth control. If it is possible to allow some parts of society to address the question of making safe, effective, and voluntary birth control available while other parts of society work to address child survival rates and literacy and education rates, we must not prolong the argument. We simply must move forward as a world with a new common understanding that we do have the capacity to create throughout this world the conditions that will lead to stabilizing population all over this world.

Those of us in the industrial countries must also listen to those in developing countries who say, "Wait a minute, what about the fact that each and every individual in an industrial country is responsible on a per person basis for twenty or thirty times as much pollution as an individual in developing countries?" In order for the world as a whole to respond to the challenge posed by the population explosion, we must also address the problems of waste and wasteful consumption in the industrial world.

The second cause of this transformed relationship between civilization and the earth is the scientific and technological revolution. We have had a difficult time coming to grips with the fantastic new powers that result from our intellectual achievements as a civilization. Remember the story of the sorcerer's apprentice who learned the magical incantation that began a process which he did not know how to stop? We have been faced with a similar challenge. Just as nuclear weapons transformed the consequences of all-out warfare, so thousands of new technologies that magnify our ability to exploit the earth have transformed the consequences of all-out exploitation.

At the dawning of the nuclear age, Albert Einstein said, "Everything has changed, except our way of thinking." Over the last few decades, human beings have tried to come to grips with a new way of thinking about warfare. We have not fully acknowledged the debt that all humankind owes to Brazil and Argentina and other nations in South America who had the courage to say, "We will abandon this technology of nuclear weapons, and we will make a pact to stop the march that other nations began toward nuclear weaponry." We have had achievements in this area.

Now we face the necessity of arriving at a new way of thinking about our relationship to technologies for exploiting the earth. New policies regarding chlorofluorocarbons (CFCs) give us a cause for hope because agreements were reached at the Earth Summit by most nations of the world that will eventually result in their phaseout. We

must build on that success by recognizing the partnership between north and south, industrial countries and developing countries. These partnerships make possible assistance from the north to those nations in the south that need access to new technologies that allow for quicker abandonment of CFCs.

Of course, some of the technologies now contributing to the destruction of the earth's environment will be more difficult to change. The internal combustion engine and all of the engines that burn fossil fuels are now putting a quantity of carbon dioxide and other greenhouse gases into the atmosphere that threatens the climate balance of the earth. We must create a new partnership on a global basis to dramatically accelerate the development of new technologies that allow and foster economic progress without environmental destruction.

Developing countries have expressed their concern about the terms of technology transfer. They fear economic exploitation. Those from developing countries should listen to those in the industrial world who say, "The fruits of invention must be safeguarded sufficiently to reward those who come up with the new inventions." This principle will encourage the flood of new discoveries upon which all humankind will rely during the transition to environmentally sound technologies.

After population and technology, the third and final cause of this transformed relationship is the most subtle but the most important. It is our way of thinking about the relationship itself. Specifically, it is our assumption that we are somehow separate from nature—that we have the right to exploit nature without any concern for the consequences of what we do. It is an assumption that we who are alive today somehow have the right to exploit the earth without considering those who still come after us in the next generation and the ones after that. This way of thinking has led to horrendous exploitation, and it must now change.

In order to change it, we must continue a dialogue among the peoples of every nation, between the peoples of industrial cultures and traditional cultures, drawing upon the wisdom that has been passed down over the thousands of generations we have been on earth. We must resist the temptation to believe that the only thing that matters is the world of thought. We must instead come to an understanding of our basic connection with the rest of nature, with each other, and with those who will come after us.

I believe the central organizing principle in the post-Cold War world must become the task of saving the earth's environment. We are close to a time when all of humankind will envision a global agenda that encompasses a kind of Global Marshall Plan, if you will, to address the causes of poverty and suffering and environmental destruction all over the earth. In order to reach that day, we must have leadership in this world. And when it does not come from heads of state, it must come from others who are close to the people

and who hear what they, especially children, are telling us.

We must gain enough momentum to overcome the two obstacles that lie in our path. The first is the obstacle of denial that prevents us from recognizing the problem as it truly exists. But even after we break through that barrier and realize the enormity of the challenge before us, we run headlong into the second obstacle, even more formidable than the first. It is called despair. How do you tell people to have hope in the face of these enormous tragedies unfolding around the world? How can we have hope when we see how large these tragedies are and how enormous these challenges have become?

We can have hope because we are human beings and because we are part of the earth. Archbishop Camara never gave up hope when democracy was snuffed out in Brazil. Nelson Mandela never gave up hope when apartheid threatened to destroy South Africa. The men and women of the formerly communist countries never gave up hope when they saw the opportunity to restore freedom in their lands. We have no right to give up hope. We must reach out to freedom, and we must understand the concept of sustainable freedom. As Martin Luther King, Jr. said, "Whenever freedom is denied to anyone, freedom is threatened for everyone."

We are one world. We do have common commitments. And we share a common hope for the future of this earth. Let us commit ourselves to one another and to the people all over this world whom we represent that we shall overcome in this great struggle.

Conceptions of International Peace and Environmental Rights: "The Remains of the Day"

BARBARA STARK*

> But increasingly we are beginning to recognize a different commonage, a common heritage not in the earth's resources but in the earth itself and in the global environment. . . . Might global danger require a new conception of commonage, one that supports international regulation that is not only extraterritorial, one that cannot wait on universal enlightenment to bring universal consent to what may be essential?
>
> Louis Henkin[1]

INTRODUCTION

The international legal system is reinventing itself in the aftermath of the Cold War.[2] The United States is in the process of shaping a

---

* Associate Professor, University of Tennessee College of Law. LL.M. 1989, Columbia University; J.D. 1976, New York University; B.A. 1974, Cornell University. I am deeply grateful to Fran Ansley, David Bederman, Mary Ellen O'Connell, and Glenn Reynolds for very helpful criticism on very short notice; to Grayfred Gray, Dean Rivkin, and Alan Thorndike for their valuable insights; and to Mart Cizek for superb research assistance. I also acknowledge the generous support of the American Council of Learned Societies, the American Philosophical Society, and the Faculty Development Program of the University of Tennessee, and the assistance of Rahul Kale of the United Nations Center for Human Rights (New York) and Alexandre Tikhonov, Secretary of the United Nations Committee for Economic, Social and Cultural Rights (Geneva).

1. Louis Henkin, *International Law: Politics, Values and Functions* 216 RECUIL DES COURS 1, 348 (1989).

2. Symposium, *After the Cold War: International Law in Transition*, 32 HARV. INT'L L.J. 321 (1991). For an account of the collapse of the Soviet Union see Serge Schmemann, *Declaring Death of Soviet Union, Russia and 2 Republics Form New Commonwealth*, N.Y. TIMES, Dec. 9, 1991, at A1. For thoughtful analyses, see Graham Allison & Robert Blackwill, *America's Stake in the Soviet Future*, 70 FOREIGN AFF., Summer 1991, at 77; Seweryn Bialer, *The Death of Soviet Communism*, 70 FOREIGN AFF., Winter 1991-1992, at 166. *See generally* ROBERT CULLEN, TWILIGHT OF EMPIRE: INSIDE THE CRUMBLING SOVIET BLOC (1991); VLADISLAV KRASNOV, RUSSIA BEYOND COMMUNISM (1991).

The end of the Cold War permitted the unprecedented coalition during the Gulf War. *See* Thomas M. Franck & Faiza Patel, *UN Police Action in Lieu of War: "The Old Order Changeth,"* 85 AM. J. INT'L. L. 63 (1991); Alvin Z. Rubinstein, *New World Order or Hollow Victory?*, 70 FOREIGN AFF., Fall 1991, at 53; Oscar Schachter, *United Nations Law in the Gulf Conflict*, 85 AM. J. INT'L. L.

new leadership role for itself in the emerging world order. The International Covenant on Civil and Political Rights,[3] in limbo since it was signed by President Carter in 1977,[4] was ratified in April 1992.[5] In a similar breakthrough, the United States signed the Charter of Paris for a New Europe in 1990, expressly affirming in an international instrument that "every individual has the right . . . to enjoy his economic, social and cultural rights."[6]

Are we ready to renew and expand our commitment to international human rights?[7] Will we focus instead on the role of "world policeman," working with and through the United Nations to maintain peace, or at least to contain conflict?[8] We might even try to

---

452 (1991); Symposium, *The Iraqi Crisis: Legal and Socio-Economic Dimensions*, 15 S. ILL. U. L.J. 411 (1991).

A separate development, the economic unification of Europe, also has implications for the international system. *See, e.g.*, Council Directive 88/361 of 24 June 1988 for the Implementation of Article 67 of the Treaty, 1988 O.J. (L 178) 5 (Free Movement of Capital Directive); Council Decision 88/591, 1988 O.J. (L 319) 1 (establishing a court of First Instance of the European Communities); John T. Lang, *The Development of European Community Constitutional Law*, 25 INT'L LAW. 455 (1991); Alan Riding, *Europeans Agree on a Pact Forging New Political Ties and Integrating Economies*, N.Y. TIMES, Dec. 11, 1991, at A1. *See generally*, JOHN PINDER, EUROPEAN COMMUNITY: THE BUILDING OF A UNION (1991); ALPO M. RUSI, AFTER THE COLD WAR: EUROPE'S NEW POLITICAL ARCHITECTURE (1991); RENÉ SCHWOK, U.S.-EC RELATIONS IN THE POST COLD WAR ERA (1991); Gregory F. Treverton, *The New Europe*, 71 FOREIGN AFF., America and the World 1991-1992, at 94.

3. International Covenant on Civil and Political Rights, *opened for signature* Dec. 19, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976); *see infra* notes 54, 58-61 and accompanying text.

4. *President Carter Signs Covenants on Human Rights*, 77 DEP'T ST. BULL. 587 (1977).

5. 138 CONG. REC. S4781-84 (daily ed. Apr. 2, 1992)(Senate Executive Session on the International Covenant on Civil and Political Rights).

6. Charter of Paris for A New Europe, Nov. 21, 1990, 30 I.L.M. 190. As Professor Damrosch observed, instruments like the United Nations Charter are generally considered "political" rather than "legal" undertakings. Lori Fisler Damrosch, *International Human Rights Law in Soviet and American Courts*, 100 YALE L.J. 2315, 2319 (1991).

7. *See* Jimmy Carter, *Keynote: The United States and the Advancement of Human Rights Around the World*, 40 EMORY L.J. 723 (1991); *cf.* Brenda Cossman, *Reform, Revolution or Retrenchment? International Human Rights in the Post-Cold War Era*, 32 HARV. INT'L L.J. 339 (1991)(considering possibility of renewed commitment to human rights on the international level).

Such a commitment necessarily implies an endorsement of the values which shape international human rights, particularly a deep respect for "human dignity." For a comprehensive and richly contextualized description of human dignity, *see* MYRES S. MCDOUGAL ET AL., HUMAN RIGHTS AND WORLD PUBLIC ORDER 367-449 (1980).

8. *See generally* Symposium, *The Use of Force in the Post-Cold War Era*, 20 DENV. J. INT'L L. & POL'Y 1 (1991); Louis Henkin, *Law and War After the Cold War*, 15 MD. J. INT'L L. & TRADE 147 (1991); Nicholas Rostow, *The*

pull up the drawbridge. It may be argued that our own survival—or at least our own standard of living[9]—is too much at risk to take on global problems, especially the overwhelming social and economic problems of the Third World.

There are difficult choices to be made and the absence of a clear consensus begs for fresh perspectives. As the United States considers the awesome responsibilities of world leadership, we would do well to ask ourselves where it is we hope to lead—what kind of world and what kind of future do we want for ourselves and for our children? It is time to reflect, to question, and to rigorously examine the normative underpinnings of the international system.[10]

After World War II, the world powers signed the United Nations Charter and the Universal Declaration of Human Rights. The two stand as an unequivocal denunciation of the atrocities of the war—a resounding, "Never again!" Never again would the world permit such a war; never again such violations of human rights. Peace and human rights are at the very foundation of modern international law.

The historical linkage between human rights and peace is obvious, but the conceptual as well as the practical linkage between them remains unclear.[11] This Article is a preliminary exploration of the

---

International Use of Force After the Cold War, 32 HARV. INT'L L.J. 411 (1991). Some commentators have argued that this might include intervention to support struggling democracies. For concise summaries of those arguments, and their refutations, by the leading international scholars in the field, see LAW AND FORCE IN THE NEW INTERNATIONAL ORDER 111-223 (Lori Fisler Damrosch & David J. Scheffer eds., 1991).

9. *See* Alan Tonelson, *What is the National Interest?*, THE ATLANTIC MONTHLY, July 1991, at 35, 37 ("Internationalism . . . has led directly to the primacy of foreign policy in American life and to the consequent neglect of domestic problems . . . .").

10. *See* Thomas M. Franck, *United Nations Based Prospects for a New Global Order*, 22 N.Y.U.J. INT'L L. & POL. 601, 603 (1990). Franck compares this post-Cold War moment to 1787, when, as Professor David Richards observed, a "'sense of challenge and opportunity fired the founders to initiate with the American people a great collective democratic deliberation on constitutionalism. . . .' Such a great collective democratic deliberation should now be going on in the world." *Id.* at 603 (quoting DAVID A.J. RICHARDS, FOUNDATIONS OF AMERICAN CONSTITUTIONALISM 20 (1989)). Similarly, Mark Janis has stated, "Now is the time to . . . modify [international law's] contemporary conceptions, both in its basic theory and in its practice." Mark W. Janis, *International Law?*, 32 HARV. INT'L L.J. 363, 364 (1991).

11. As Professor Sohn pointed out, "It is an axiom that there is a connection between [them], but it has not been investigated in depth." Letter from Louis B. Sohn, Professor of Law, University of Georgia (Jan. 8, 1991) (on file with the author). The linkage is often noted in passing. As Professor Schachter recently observed, for example, "economic and social deficiencies . . . contribute to internal tensions and to interstate conflict." Schachter, *supra* note 2, at 473.

Efforts to understand this relationship may be increasing. *See, e.g.*, Jimmy Carter, *The Greatest Human Rights Crime: War*, 13 HAMLINE L. REV. 469 (1990).

relationship between our conceptions of peace and human rights, using environmental rights[12] as a case study. My basic thesis is that there is a fundamental tension between "peace" and "human rights"—a tension deriving less from any inherent normative conflict than from their respective (and complementary) spheres of influence in the political world. Since "peace" addresses the conduct of states in their external relations with other states, it is necessarily governed by universal standards, i.e., standards established and shared by all sovereign states. Human rights law,[13] in contrast, focuses on the conduct of states towards their own people. Although human rights norms are framed in terms of universal standards, domestic enforcement is shaped by local culture and circumstances.[14]

For discussions of "humanitarian intervention," or the use of force to protect human rights, see, for example, FERNANDO R. TESÓN, HUMANITARIAN INTERVENTION: AN INQUIRY INTO LAW AND MORALITY (1988); Anthony D'Amato, *The Invasion of Panama Was a Lawful Response to Tyranny*, 84 AM. J. INT'L. L. 516 (1990); Lori Fisler Damrosch, *Commentary on Collective Military Intervention to Enforce Human Rights*, *in* LAW AND FORCE IN THE NEW INTERNATIONAL ORDER, *supra* note 8, at 215; Tom J. Farer, *An Inquiry into the Legitimacy of Humanitarian Intervention*, *in* LAW AND FORCE IN THE NEW INTERNATIONAL ORDER, *supra* note 8, at 185; Vladimir Kartashkin, *Human Rights and Humanitarian Intervention*, *in* LAW AND FORCE IN THE NEW INTERNATIONAL ORDER, *supra* note 8, at 202; James A.R. Nafziger, *Self-determination and Humanitarian Intervention in a Community of Power*, 20 DENV. J. INT'L L. & POL'Y 9 (1991). For some preliminary thoughts on the linkage between economic rights and peace, see Barbara Stark, *Nurturing Rights: An Essay on Women, Peace, and International Human Rights*, 13 MICH. J. INT'L. L. 144 (1991).

Professors MacDougal and Chen and the late Professor Lasswell have argued in a vast body of work that these conceptions are necessarily intertwined, both with each other and with other constituent dimensions of "world order." *See, e.g.*, MCDOUGAL ET AL., *supra* note 7.

Some international instruments explicitly establish a linkage. The 1977 Protocols to the Geneva Convention, for example, prohibit "methods or means of warfare which are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment." Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), June 8, 1977, art. 35, U.N.T.S. 3, 21 (entered into force Dec. 7, 1978).

For a graphic account of the impact of the Persian Gulf War on the environment, see Nicholas A. Robinson, *International Law and the Destruction of Nature in the Gulf War*, 21 ENVTL. POL'Y & L. 216 (1991).

12. Environmental rights often have been referred to as "third generation" human rights. In the now classic formulation, "first generation" civil and political rights correspond to the French Revolution's "liberté," "second generation" economic and social rights correspond to "égalité," and "third generation" collective rights to "fraternité" or "solidarité." Stephen P. Marks, *Emerging Human Rights: A New Generation for the 1980s?*, 33 RUTGERS L. REV. 435, 441 (1981).

13. As used in this Article, "human rights" refers to the familiar civil, political, economic, social, and cultural rights set forth in the International Bill of Rights. *See infra* notes 54-55 and text accompanying notes 50-72.

14. I am not suggesting that domestic adulteration is acceptable practice

Environmental rights partake of both regimes, although the case for considering them "human rights"[15] is probably stronger. Like other human rights, environmental rights have been tailored to local conditions and require intrastate enforcement. Like peace, however, environmental rights demand agreed upon norms and effective interstate implementation. To be fully realized, environmental rights require not only considerable coordination, but fundamental coherence between the interstate and intrastate regimes, a coordination and a coherence that has never before been achieved or even attempted.[16]

This Article compares the pressures that have produced international regimes dealing with peace, human rights in general, and environmental rights in particular. It also considers the concessions each regime has had to make to state sovereignty.[17] Why have

---

under the human rights treaties. Many of the states that are parties to those treaties file reservations to the "universal standards" or objections to the reservations of other states. *See, e.g.*, *infra* notes 127-28. Nor is there scholarly agreement as to the range of permissible variation. *See generally* Jack Donnelly, *Cultural Relativism and Universal Human Rights*, 6 HUM. RTS. Q. 400 (1984)(considering competing claims of relativism and universalism); Alison D. Renteln, *The Unanswered Challenge of Relativism and the Consequences for Human Rights*, 7 HUM. RTS. Q. 514 (1985)(pointing out common assumptions and pitfalls of universalism). As a practical matter, however, it is generally recognized that states are loathe to judge others, lest they be judged. *See infra* note 74.

In Barcelona Traction, Light and Power Company, Ltd. (Belg. v. Spain), 1970 I.C.J. 3, 32 (Feb. 5), the International Court of Justice noted that "all States . . . have a legal interest in [the protection of] . . . the basic rights of the human person, including protection from slavery and racial discrimination" and held that all states accordingly could bring a claim when "obligations *erga omnes*" were violated. To date, however, no state has sought to rely on *Barcelona Traction* for this principle. LOUIS HENKIN ET AL., INTERNATIONAL LAW: CASES AND MATERIALS 531 n.2 (2d ed. 1987) [hereinafter HENKIN ET AL., INTERNATIONAL LAW]; *see also infra* note 33.

15. For a brief history of efforts to establish environmental rights as human rights, see Melissa Thorme, *Establishing Environment As a Human Right*, 19 DENV. J. INT'L L. & POL'Y 301, 303-05 (1991); *see also infra* text accompanying notes 110-12 (discussing human rights and the environment); *cf. infra* text accompanying notes 94-95 (discussing environment as a security issue).

16. Whether peace and other human rights similarly require such coordination and coherence to be fully realized is a more profound question that is beyond the scope of this Article.

As Sir Geoffrey Palmer hopefully observed, "The extraordinary changes in world order that have recently taken place must surely increase the chances of achieving change in the methods of making international environmental law." Geoffrey Palmer, *New Ways to Make International Environmental Law*, 86 AM. J. INT'L. L. 259, 259 (1992).

17. As the International Court of Justice pointed out in Military and Paramilitary Activities (Nicar. v. U.S.), 1986 I.C.J. 14 (June 27), "the whole of international law rests [on the fundamental principle of state sovereignty]." *See generally* Walter F. Mondale, *Human Rights and the Environment: Facing a New World Order*, 16 VT. L. REV. 449 (1992).

sovereign states, often at considerable risk and expense, pledged to promote peace, human rights, and environmental rights? What explains the sustained commitment to norms once dismissed as hortatory and now recognized as enforceable[18] law?[19] Equally important, what are the concessions, or accommodations, to state interests that have been necessary to achieve this consensus? My purpose here is not to undertake a comprehensive study, but to provide the general reader with both a conceptual overview of these three distinct but

---

18. Those skeptics who still doubt the "enforceability" of international law might recall Louis Henkin's oft-quoted observation that "almost all nations observe almost all principles of international law and almost all of their obligations almost all of the time." LOUIS HENKIN, HOW NATIONS BEHAVE 47 (2d ed. 1979)(emphasis omitted). In addition to the "horizontal sanctions" arising from a shared and fundamental, albeit imperfect, respect for international law, enforcement may be obtained through the International Court of Justice. *See* Mary Ellen O'Connell, *The Prospects for Enforcing Monetary Judgments of the International Court of Justice: A Study of Nicaragua's Judgment Against the United States*, 30 VA. J. INT'L L. 891 (1990). *But see* Jonathan I. Charney, *Disputes Implicating the Institutional Credibility of the Court: Problems of Non-Appearance, Non-Participation and Non-Performance*, *in* THE INTERNATIONAL COURT OF JUSTICE AT A CROSSROADS 288, 293-97 (Lori Fisler Damrosch ed., 1987)(citing cases showing decline in states' willingness to accept the authority of the court). The United Nations Security Council also enforces international law. *See, e.g.*, S.C. Res. 687, U.N. SCOR, 44th Sess., 2981st mtg., U.N. Doc. S/RES/687 (1991). For conceptual descriptions of "enforcers" in international law and their limitations, see W. Michael Reisman, *Sanctions and Enforcement*, *in* MYRES S. MCDOUGAL & W. MICHAEL REISMAN, INTERNATIONAL LAW ESSAYS 381, 413-20 (1981).

It is generally recognized that domestic courts remain the most important enforcers of international law. *See infra* note 88. For an overview of the ways in which human rights law has already affected domestic jurisprudence, see Kathryn Burke et al., *Application of International Human Rights Law in State and Federal Courts*, 18 TEX. INT'L L.J. 291 (1983). Some commentators have suggested future possibilities for domestic enforcement of international human rights law. *See, e.g.*, Farooq Hassan, *The Doctrine of Incorporation: New Vistas for the Enforcement of International Human Rights?*, 5 HUM. RTS. Q. 68, 69 (1983)(suggesting that the Tenth Circuit's reliance on international norms in Rodriguez-Fernandez v. Wilkinson, 654 F.2d 1382 (10th Cir. 1981), "usher[ed] in new vistas for the domestic protection of internationally recognized human rights" through incorporation into federal common law); Alan Brudner, *The Domestic Enforcement of International Covenants on Human Rights: A Theoretical Framework*, 35 U. TORONTO L.J. 219, 233 (1985). *But see* Linda Greenhouse, *High Court Backs Seizing Foreigner for Trial in U.S.*, N.Y. TIMES, June 16, 1992, at A1 (Mexican citizen, kidnapped by the United States government in what Justice Stevens characterizes as "a flagrant violation of international law," may nonetheless be tried in the United States) (quoting United States v. Alvarez-Machain, 112 S. Ct. 2188, 2203 (1992) (Stevens, J., dissenting)).

19. This development is often discussed in terms of a progression from "soft" to "hard" law. Adherence to various hortatory declarations has historically been obtained through such a process. The Treaty on Principles Governing States in the Exploration and Use of Outer Space, Jan. 27, 1967, 18 U.S.T. 2410, 610 U.N.T.S. 205, for example, originated in "soft" accession to general principles. *See generally*, C.M. Chinkin, *The Challenge of Soft Law: Development and Change in International Law*, 38 INT'L & COMP. L.Q. 850 (1989).

related regimes and an analytic framework in which to compare them. I conclude that we must recognize and transcend the limitations of the peace and human rights regimes if we are to develop effective international environmental law.

I. PEACE[20]

A. Impetus for Acceptance

The prohibition of the use of force in the United Nations Charter represents a substantial limitation on traditional conceptions of state sovereignty. International law has historically been understood as the law of nation states—the rules agreed upon by the nations of the world to govern their dealings with one another. Under international law each state is recognized as autonomous and sovereign; none is subject to the authority of another. The only recognized limitations on state sovereignty were those to which the state itself acquiesced, either explicitly in a treaty or through consistent custom.

Historically, waging war was a sovereign prerogative.[21] It might be disapproved, it might have tremendous political costs, but it was

---

20. For present purposes, "peace" may be understood as it is used in the *Declaration on the Right of Peoples to Peace*, G.A. Res. 11, U.N. GAOR, 39th Sess., Supp. No. 51, at 22, U.N. Doc. A/39/51 (1985), which provides in pertinent part that the right to peace "demands that the policies of States be directed towards the elimination of the threat of war, particularly nuclear war, the renunciation of the use of force in international relations and the settlement of international disputes by peaceful means on the basis of the Charter of the United Nations." Article I of the United Nations Charter provides that the purposes of the United Nations include:

> 1. To maintain international peace and security, and to that end: to take effective collective measures for the prevention and removal of threats to the peace, and for the suppression of acts of aggression or other breaches of the peace, and to bring about by peaceful means, and in conformity with the principles of justice and international law, adjustment or settlement of international disputes or situations which might lead to a breach of the peace;
>
> 2. To develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples, and to take other appropriate measures to strengthen universal peace.

U.N. CHARTER art. 1.

This definition is concededly rather arid and legalistic. For those craving more philosophical treatments, see, for example, HANNAH ARENDT, THE HUMAN CONDITION (1958); NICCOLO MACHIAVELLI, THE DISCOURSES (Max Lerner ed., 1950); THE REPUBLIC OF PLATO (Allan Bloom trans., Basic Books 1968).

21. IAN BROWNLIE, INTERNATIONAL LAW AND THE USE OF FORCE 14-50 (1981). For a description of "just war" prior to the Peace of Westphalia in 1648, see MALCOLM SHAW, INTERNATIONAL LAW 539-41 (1986). War could be—and was—declared to seize territory, to redress an insult, for the glory of God, or merely on a whim. *See generally* FREDERICK H. RUSSELL, THE JUST WAR IN THE MIDDLE AGES (1975); MICHAEL WALZER, JUST AND UNJUST WARS (1977).

not generally considered *illegal* under international law until the Kellogg-Briand Pact in 1928.[22] Even with the recognition of a legal prohibition against war, it was another twenty years before legal structures or mechanisms for averting war were devised.[23] It took the devastation of World War II to convince the world powers that limits had to be imposed for the security, even the survival, of all states in a nuclear age. "Determined to save succeeding generations from the scourge of war, which twice in our lifetime has brought untold sorrow to mankind,"[24] nation states banded together to prohibit the "threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."[25] The United Nations Charter represented a significant concession on the part of sovereign states. It also represented their collective acknowledgement of a preemptive common objective—mutual survival. The urge to punish a competitive neighbor, to seize particularly attractive territory, or even to protect the state from a real political or economic threat was not worth the risk of annihilation. As Professor Henkin has explained, "Peace was the paramount value. . . . Peace was more important than progress and more important than justice."[26]

B. Concessions to State Sovereignty

While the United Nations Charter's prohibition on the use of force represents a major restraint on hitherto unfettered sovereign

---

22. General Treaty for the Renunciation of War, Aug. 27, 1928, 46 Stat. 2343, 94 L.N.T.S. 57.

> Article I
>
> The High Contracting Parties solemnly declare in the names of their respective peoples that they condemn recourse to war for the solution of international controversies, and renounce it as an instrument of national policy in their relations with one another.
>
> Article II
>
> The High Contracting Parties agree that the settlement or solution of all disputes or conflicts of whatever nature or of whatever origin they may be, which may arise among them, shall never be sought except by pacific means.

Id. at 2345-46, 94 L.N.T.S. at 63.

23. The Covenant of the League of Nations provided for resort to "pacific settlement" before going to war. *See* LEAGUE OF NATIONS COVENANT arts. 12-16. While this scheme was intended to deter war, it did not purport to enforce a legal prohibition against it: "War, as such, was not made illegal but only where begun without complying with the requirement." HENKIN ET AL., INTERNATIONAL LAW, *supra* note 14 at 668-70 (quoting D.W. BOWETT, THE LAW OF INTERNATIONAL INSTITUTIONS 15-16 (1963)).

24. U.N. CHARTER pmbl.

25. U.N. CHARTER art. 2, para. 4.

26. Louis Henkin, *Use of Force: Law and U.S. Policy*, *in* LOUIS HENKIN ET AL., RIGHT V. MIGHT: INTERNATIONAL LAW AND THE USE OF FORCE 37, 38-39 (1989).

states, two important limitations make this prohibition acceptable. First, carefully crafted exceptions permit the state to use force if necessary for its own self-defense or at the request of a friendly state which has been attacked ("collective self-defense").[27] Thus, the use of force in response to a prior *unlawful* use of force is permitted under the Charter.[28]

Second, and equally important, the prohibition only applies to the interstate use of force. Article 2(4) of the Charter by its terms only restricts armed conflict between different sovereign states.[29] Civil wars,[30] even the harshest suppression of domestic insurgents or minority populations, are regarded more as internal matters than as appropriate subjects of international intervention.[31] Sovereign states retain their right to use force internally to protect themselves against "domestic" threats to their own security. The notable exception, of course, is that states cannot use force if in doing so they violate human rights.[32] The international community, however, remains reluctant to intervene in—or even criticize—another state's "internal policies."[33]

---

27. U.N. CHARTER art. 51.

28. 2 RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW, § 905 (1987) (Unilateral Remedies); *see also* David R. Penna, *The Right to Self-Defense in the Post-Cold War Era: The Role of the United Nations*, 20 DENV. J. INT'L L. & POL'Y 41 (1991).

29. *See also Declaration on the Inadmissibility of Intervention into the Domestic Affairs of States*, G.A. Res. 2131, U.N. GAOR, 20th Sess., Supp. No. 14, at 11, U.N. Doc. A/6014 (1965).

30. *See* Mary Ellen O'Connell, *Continuing Limits on UN Intervention in Civil War*, 67 IND. L.J. 903 (1992); Oscar Schachter, *The Right of States to Use Armed Force*, 82 MICH. L. REV. 1620, 1641-45 (1984).

31. *See* Louis Henkin, *The Invasion of Panama Under International Law: A Gross Violation*, 29 COLUM. J. TRANSNAT'L L. 293 (1991); Damrosch, *supra* note 11; Kartashkin, *supra* note 11; Abraham D. Sofaer, *The Legality of the United States Action in Panama*, 29 COLUM. J. TRANSNAT'L L. 281 (1991).

32. The Security Council is considering appropriate limits on the right to use force internally in the case of Yugoslavia. *See* O'Connell, *supra* note 30, at 909-12. The United Nations recently adopted trade sanctions against Yugoslavia in an effort to promote peace in Bosnia and Herzegovina. Paul Lewis, *U.N. Votes 13-0 for Embargo on Trade with Yugoslavia; Air Travel and Oil Curbed*, N.Y. TIMES, May 31, 1992, § 1, at 1; *Excerpts From U.N. Resolution: "Deny Permission,"* N.Y. TIMES, May 31, 1992 § 1, at 8.

33. [Some argue that] one State should be free to invade another country to prevent a holocaust or to depose a genocidal regime. That argument is seductive but specious. . . . In fact, no State has pressed for exception to the law of the Charter that would permit invading another country to remedy even the grossest of human rights violations. In fact, no State has gone to war against another State for the purpose of ending human rights violations.

Henkin, *supra* note 8, at 156.

C. Compliance

The international peacekeeping regime focuses on the acts of states, that is, acts of aggression by one state against another.[34] While acts of individuals may well have an impact on peace, they neither trigger sanctions nor justify the use of force by the target state, unless they can be positively attributed to a state.[35] Terrorism, for example, has been discouraged in a series of multilateral treaties,[36] but it generally does not give the target state the right to use force in self-defense.[37]

How does the international system deal with violating states?[38] Article 51 self-defense, described above,[39] is basically an interim measure under the United Nations Charter scheme. A state may utilize self-defense "*until* the Security Council has taken the measures necessary to maintain international peace and security."[40] Under Article 41, the Security Council has a broad range of options, including resolutions of condemnation, economic sanctions, "complete or partial interruption of economic relations and . . . communication, and the severance of diplomatic relations."[41] Moreover, the Security Council "may take such action by air, sea, or land forces as may be necessary to maintain or restore international peace and security," including the use of force, should it decide that Article 41 measures "would be inadequate or have proved to be inadequate."[42] It remains an open question whether Security Council action preempts further independent action by the target state or its allies.[43]

---

34. *Definition of Aggression Resolution* art. 1, G.A. Res. 3314, U.N. GAOR, 29th Sess., Supp. No. 31, at 142, 143, U.N. Doc. A/9631 (1975).

35. Military and Paramilitary Activities, *supra* note 17, at 354-63.

36. *See, e.g.*, Convention for the Suppression of Unlawful Seizure of Aircraft, Dec. 16, 1970, 22 U.S.T. 1641, 10 I.L.M. 133 (Hague Convention); Convention for the Suppression of Unlawful Acts Against the Safety of Civil Aviation, Sept. 23, 1971, 24 U.S.T. 565, 10 I.L.M. 1151 (Montreal Convention).

37. A state may be held accountable for terrorist acts under norms governing state responsibility if the state subsidized, supported, or otherwise affirmatively encouraged terrorists. Richard B. Lillich & John M. Paxman, *State Responsibility for Injuries to Aliens Occasioned by Terrorist Activities*, 26 AM. U.L. REV. 217, 307-09 (1977); *cf.* Geoffrey M. Levitt, *Intervention to Combat Terrorism and Drug Trafficking*, *in* LAW AND FORCE IN THE NEW INTERNATIONAL ORDER, *supra* note 8, at 224, 227 (the United States bombing of Libya was carefully justified by focusing on "a pattern of incidents which, taken as a whole, amounts to an ongoing armed aggression" likely to continue absent "preemptive" action by the United States).

38. *See* U.N. CHARTER art. 51.

39. *See supra* notes 27-28 and accompanying text.

40. U.N. CHARTER art. 51 (emphasis added).

41. U.N. CHARTER art. 41.

42. U.N. CHARTER art. 42.

43. Henkin would leave this to the Security Council. Henkin, *supra* note 8, at 161. There is also the question of the extent of Security Council authority over

The Article 43 regime, calling for "special agreements" under which member states would provide armed forces for international security, has never been implemented because of the stalemate in the Security Council during the Cold War.[44] While the scope and usefulness of the Article 43 regime remains to be proved, two points are critical for present purposes. First, the Charter's peacekeeping regime contemplates an international, collective response to the use of force. Second, until and unless the member states commit substantial resources to Article 43 forces[45]—or an alternative is devised and adopted—international peacekeeping is structured more to respond to actual acts of aggression than to defuse simmering hostilities and prevent their outbreak.

While some commentators have argued that the prohibition against the use of force has been selectively enforced, on one has claimed that it is a variable norm. This is not to say that the use of force is unambiguous under the Charter. Louis Henkin takes the position that there may be absolutely no transboundary use of force except in the case of self-defense against armed attack.[46] Oscar Schachter has explained that the use of force under Article 2(4) must be "proportional."[47] And Anthony D'Amato has argued that under the doctrine of humanitarian intervention, the limits on the legitimate use of force are even more liberal than those accepted in the Persian Gulf War.[48] All of these authors, however, like most international

---

member states. During the Persian Gulf War, commentators noted that Security Council Resolution 678 imposed no legal obligation on the United States to use armed force. *See, e.g.*, Michael J. Glennon, *The Constitution and Chapter VII of the United Nations Charter*, 85 AM. J. INT'L L. 74 (1991). *See generally* S.C. Res. 678, U.N. SCOR, 45th Sess., Resolutions and Decisions of 1990 at 27, U.N. Doc. S/INF/46 (1991).

44. *But see Uniting for Peace Resolution*, G.A. Res. 377A, U.N. GAOR, 5th Sess., Supp. No. 20, at 10, U.N. Doc. A/1775 (1951)(urging the Security Council to promote the formation of special agreements under Article 43). This resolution was passed during the Korean War and is usually considered an exceptional case. For a concise summary, see DAVID J. SCHEFFER, UNITED NATIONS ASS'N OF THE UNITED STATES OF AMERICA, THE UNITED NATIONS IN THE GULF CRISIS AND OPTIONS FOR U.S. POLICY 8 (Occasional Papers 1991).

Franck and Patel say that the lack of Article 43 military forces does not matter. Franck & Patel, *supra* note 2, at 66. Many other scholars disagree. *See, e.g.*, Glennon, *supra* note 40.

45. For a recent assessment, see Mary Ellen O'Connell, *Enforcing the Prohibition on the Use of Force: The U.N.'s Response to Iraq's Invasion of Kuwait*, 15 S. ILL. U. L.J. 453, 482-84 (1991).

46. Henkin, *supra* note 26, at 37, 44-45.

47. Oscar Schachter, *In Defense of International Rules on the Use of Force*, 53 U. CHI. L. REV. 113, 120 (1986). For an updated analysis, see O'Connell, *supra* note 45, at 481-86.

48. Anthony D'Amato, Book Review, 85 AM. J. INT'L L. 201, 202 (1991)(reviewing HENKIN ET AL., RIGHT V. MIGHT: INTERNATIONAL LAW AND THE USE OF FORCE (1989)); *see also* Tesón, *supra* note 11; D'Amato, *Invasion of Panama*, *supra* note 11.

scholars who analyze the use of force, seek to articulate a standard to be applied universally, in all situations.[49] All states, obviously, have an interest in a clear standard, fairly applied.

II. HUMAN RIGHTS

A. Impetus for Acceptance

Just as the United Nations Charter's restrictions on the use of force limit a state's options in its dealings with other states, the growing body of international human rights law restricts a state's options in its dealings with its own people.[50] A series of international instruments requires states to respect and protect what the Charter refers to as "the dignity and worth of the human person."[51] The Universal Declaration of Human Rights[52] fleshes out this concept, establishing the framework for the "International Bill of Rights."[53] This consists of the International Covenant on Civil and Political Rights[54] (Political Covenant or ICCPR) and the International Covenant on Economic, Social and Cultural Rights[55] (Economic Covenant or ICESCR).

---

49. For a summary of the impact of United Nations peacekeeping from 1945-1984, see Eric Stein, *The United Nations and the Enforcement of Peace*, 10 MICH. J. INT'L L. 304, 314 (1989) (quoting Ernst B. Haas, *The Collective Management of International Conflict, 1945-1984*, *in* U.N. INST. FOR TRAINING AND RESEARCH, THE UNITED NATIONS AND THE MAINTENANCE OF INTERNATIONAL PEACE AND SECURITY at 3, 19, U.N. Sales No. E.87.III.K.ST/20 (1987)).

50. *See generally* Tom J. Farer, *Human Rights in Law's Empire: The Jurisprudence War*, 85 AM. J. INT'L L. 117 (1991); Rosalyn Higgins, *Conceptual Thinking About the Individual in International Law*, 24 N.Y.L. SCH. L. REV. 11 (1978). The restrictions may also limit a state's options in its dealings with foreign nationals and refugees, i.e., stateless people. *See* Arthur C. Helton, *The Mandate of U.S. Courts to Protect Aliens and Refugees Under International Human Rights Law*, 100 YALE L.J. 2335 (1991).

51. U.N. CHARTER pmbl.; *see also* JAMES W. NICKEL, MAKING SENSE OF HUMAN RIGHTS 51-52 (1987)(arguing that "minimally good lives" should be the focus of human rights).

52. G.A. Res. 217(A) (1948). The Universal Declaration "is not in terms a treaty instrument." Secretary-General, *1971 Survey of International Law*, U.N. Doc. A/CN. 4/425, at 196.

53. *See* LOUIS HENKIN, THE AGE OF RIGHTS 16-18 (1990).

54. International Covenant on Civil and Political Rights, *opened for signature* Dec. 19, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976)[hereinafter Political Covenant]; *see also* THE INTERNATIONAL BILL OF RIGHTS: THE COVENANT ON CIVIL AND POLITICAL RIGHTS (Louis Henkin ed., 1981).

55. International Covenant on Economic, Social and Cultural Rights, *opened for signature* Dec. 16, 1966, 993 U.N.T.S. 3 (entered into force Jan. 3, 1976)[hereinafter Economic Covenant]. A group of distinguished experts in international law met in Maastricht, the Netherlands in 1986. *See The Limburg Principles on the Implementation of the International Covenant on Economic, Social and Cultural Rights*, U.N. Doc. E/CN.4/1987/17, Annex, *reprinted in* 9 HUM. RTS. Q. 121, 122 (1987). The group agreed unanimously that these principles "reflect[ed] the present state of international law" unless specifically qualified as a "recommendation." *Id.* at 121.

ICCPR addresses "first generation" civil and political rights, what Philip Alston aptly called "rights to freedom."[56] These rights are most like the negative rights familiar to those of us in the United States from our own Constitution.[57] Negative rights forbid a state from interfering with its peoples' freedoms of "thought, conscience and religion,"[58] "expression,"[59] and "liberty and security,"[60] and from denying equal protection of the law.[61] Although the United States did not ratify the Political Covenant until 1992,[62] the nation's pervasive influence and the persistent appeal of its Constitution played an important part in familiarizing the rest of the world with these rights.[63]

"Second generation" economic and social rights are set out in the Economic Covenant. Like ICCPR, the Economic Covenant is predicated on the "dignity and worth of the human person."[64] ICESCR recognizes that civil and political rights cannot be realized unless basic human needs are met.[65] Economic rights were a major concern for the former colonial Third World states who joined the United Nations in the 1960s. Article 11.1 of ICESCR, for example, provides in pertinent part: "The States Parties to the present Covenant recognize the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing, and to the continuous improvement of living condi-

---

56. Philip Alston, *A Third Generation of Solidarity Rights: Progressive Development or Obfuscation of International Human Rights Law?*, 29 NETH. INT'L L. REV. 307, 310 (1982).

57. *See, e.g.*, HURST HANNUM & RICHARD B. LILLICH, MATERIALS ON INTERNATIONAL HUMAN RIGHTS AND U.S. CONSTITUTIONAL LAW (1985); Richard B. Lillich, *The Constitution and International Human Rights*, 83 AM. J. INT'L L. 851 (1989); Richard B. Lillich & Hurst Hannum, *Linkages Between International Human Rights and U.S. Constitutional Law*, 79 AM. J. INT'L L. 158 (1985).

58. Political Covenant, *supra* note 54, art. 18, para. 1.

59. *Id.* art. 19, para. 2.

60. *Id.* art. 9, para. 1.

61. *Id.* art. 16.

62. *See supra* note 5.

63. *See* William J. Brennan, Jr., *The Worldwide Influence of the United States Constitution as a Charter of Human Rights*, 15 NOVA L. REV. 1 (1991). *See generally* authorities cited *supra* note 57.

64. Economic Covenant, *supra* note 55, pmbl.

65. *See* U.N. GAOR 3d Comm., 6th Sess., 358th-372d, 411th-417th mtgs., at 67-150, 399-499, U.N. Docs. A/C.3/SR.358-.372, .411-.417 (1951-1952)(general debates on draft international covenant on human rights). The decisions resulting from these debates are contained in the *Report of the Third Committee*, U.N. GAOR, 6th Sess., Annexes, Agenda Item 29, at 37, U.N. Doc. A/2112 (1952), and are discussed by David M. Trubeck, *Economic, Social and Cultural Rights in the Third World*, *in* HUMAN RIGHTS IN INTERNATIONAL LAW: LEGAL AND POLICY ISSUES 205, 211 n.17 (Theodor Meron ed., 1984). *See generally* Philip Alston, *U.S. Ratification of the Covenant on Economic, Social and Cultural Rights: The Need for an Entirely New Strategy*, 84 AM. J. INT'L L. 365 (1990).

tions."[66] Under Article 12.1, the parties recognize "the right of everyone to the enjoyment of the highest attainable standard of physical and mental health."[67]

Both covenants have been ratified by a substantial majority of the United Nations member states, including virtually all of the Western states and Japan.[68] The "conventional wisdom" is that two separate covenants were drafted "because of the East-West split and a disagreement over the value of socio-economic rights."[69] Some commentators attribute the division more to the differences in "the nature of the legal obligations and the systems of supervision that could be imposed."[70] Economic rights, which might require significant state expenditures, were to be achieved "progressively," while civil and political rights, which depended more on state restraint, were to be implemented immediately.[71] The interdependence of the two covenants, and the fallacy of asserting the superiority of either, are now well-established.[72]

B. Concessions to State Sovereignty

States give up some of their sovereignty by adhering to the covenants, although the extent of that relinquishment varies.[73] Two

---

66. Economic Covenant, *supra* note 55, art. 11, para. 1.

67. *Id.* art. 12, para. 1.

68. *See* BARRY E. CARTER & PHILLIP R. TRIMBLE, INTERNATIONAL LAW: SELECTED DOCUMENTS 357 n.*, 376 n.* (1989)(listing states which have ratified the ICCPR and the ICESCR).

69. David P. Forsythe, Book Review, 8 HUM. RTS. Q. 540, 541 (1986)(reviewing A. GLENN MOWER, JR., INTERNATIONAL COOPERATION FOR SOCIAL JUSTICE: GLOBAL AND REGIONAL PROTECTION OF ECONOMIC/SOCIAL RIGHTS (1985)).

70. D.J. HARRIS, CASES AND MATERIALS ON INTERNATIONAL LAW 666 n.1 (4th ed. 1991).

71. Trubeck, *supra* note 65, at 205-33.

72. *See Resolution on the Indivisibility and Interdependence of Economic, Social, Cultural, Civil and Political Rights*, G.A. Res. 130, U.N. GAOR, 44th Sess., Supp. No. 49, at 209, U.N. Doc. A/44/49 (1989)(accepted Dec. 15, 1989); Melanie Beth Oliviero, *Human Needs and Human Rights: Which Are More Fundamental?*, 40 EMORY L.J. 911 (1991); Michael W. Giles, *Comments on Oliviero Article*, 40 EMORY L.J. 939. *See generally*, Russell L. Barsh, Current Development, *A Special Session of the UN General Assembly Rethinks the Economic Rights and Duties of States*, 85 AM. J. INT'L L. 192, 199 (1991)(noting recent "linkage of human rights with the conditions for capitalism").

For a discussion of the limits of the positive-negative dichotomy in the context of ICESCR, see Philip Alston & Gerard Quinn, *The Nature and Scope of States Parties' Obligations Under International Covenant on Economic, Social and Cultural Rights*, 9 HUM. RTS. Q. 156, 159-60 (1987).

73. Sweden, for example, incurred no further obligations by ratifying ICESCR:

> Prior to ratification . . . pertinent Swedish legislation had been submitted to a careful review in order to ascertain to what extent it was in conformity with the [Economic] Covenant. No major adjustments had then been

important caveats preserve enough state autonomy to make this palatable. First, except for a handful of peremptory norms, there is little risk of interstate enforcement or even censure of human rights violations.[74]

Peremptory norms, or *jus cogens*, are norms "accepted and recognized by the international community of states as a whole as [norms] from which no derogation is permitted."[75] These norms include prohibitions against genocide, torture, racial discrimination, and apartheid.[76] Violations occur, of course, but as Professor Henkin pointed out, no state claims that torture is legal.[77] If *jus cogens* is in fact violated, the international community may impose sanctions on the offending state, as it did most notably in the case of South Africa.[78]

---

deemed necessary. Subsequent to ratification, any proposals for new legislation falling within the area covered by the Covenant must likewise be submitted to a corresponding review before their adoption as law in order to guarantee compatibility.

Report on the Second Session, U.N. Committee on Economic, Social and Cultural Rights, 2d Sess., at 26, U.N. Doc. E/1988/14, E/C.12/1988/4.

74. In response to the Chinese government's attack on demonstrating students in Tiananmen Square, for example, President Bush expressed "deep regret," and Japanese Prime Minister Sousuke Uno said he was "praying for a return to calm." *World Leaders React to Bloodshed in China*, JAPAN ECONOMIC NEWSWIRE, June 4, 1989, *available in* LEXIS, Nexis Library, JEN File; *see also* Ted Morello, *Chinese Still Welcome to Join UN Peacekeepers*, THE CHRISTIAN SCI. MONITOR, June 21, 1989, at 4; *cf. Canada Announces Measures to Protest at Chinese Crackdown*, REUTER LIB. REP., June 30, 1989, *available in* LEXIS, Nexis Library, LBYRPT File (Canada to withdraw support for three projects worth 9.1 million dollars, and also to withdraw its ambassador in Peking, but not to cut off all diplomatic and business ties "for fear of isolating China in the international community").

75. Vienna Convention on the Law of Treaties, May 23, 1969, art. 53, 1155 U.N.T.S. 331, 334 (entered into force Jan. 27, 1980). *See generally* RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW § 102 cmt. k, Reporter's note 6 (1987); Anthony D'Amato, *It's a Bird, It's a Plane, It's Jus Cogens!*, 6 CONN. J. INT'L L. 1 (1990).

76. *See* Richard B. Bilder, *An Overview of International Human Rights Law*, *in* GUIDE TO INTERNATIONAL HUMAN RIGHTS PRACTICE 3, 15-17 (Hurst Hannum ed., 1984).

77. HENKIN, *supra* note 53, at 21.

78. S.C. Res. 418, U.N. SCOR, 32d Sess., Resolutions and Decisions of 1977, at 5, U.N. Doc. S/INF/33 (1978). *See also* Douglas G. Anglin, *United Nations Economic Sanctions Against South Africa and Rhodesia*, *in* THE UTILITY OF INTERNATIONAL ECONOMIC SANCTIONS 23, 34-38 (David Leyton-Brown ed., 1987).

Victims may also seek relief in the domestic courts of other states willing to assert jurisdiction. *E.g.*, Filartiga v. Pena-Irala, 630 F.2d 876 (2d Cir. 1980)(United States court has jurisdiction over wrongful death action brought under Alien Tort Statute, 28 U.S.C. § 1350 (1988), for torture in Paraguay of plaintiffs' son); *see also* Nelson v. Saudi Arabia, 923 F.2d 1528 (11th Cir. 1991)(United States court could properly assert jurisdiction under Foreign Sovereign Immunities Act of 1976, 28 U.S.C. §§ 1602-1611 (1988 & Supp. II 1990), where engineer, recruited and hired in United States for hospital job in Saudi Arabia, was tortured by government

Consensus with respect to other human rights is problematic.[79] The process of acceding to the international covenants represents some agreement, but there is an ongoing debate as to whether it is enough.[80] The Optional Protocol,[81] which establishes an international enforcement regime, and various regional regimes,[82] which promote enforcement on the regional level, define smaller transnational communities which accept and enforce shared norms. Still, these have generated relatively scant precedent.[83] States remain reluctant to participate in regimes that allow other states to judge them. The

---

agents for reporting safety violations), *cert. granted*, 112 S. Ct. 2937 (1992). *See generally* Anne-Marie Burley, *The Alien Tort Statute and the Judiciary Act of 1789: A Badge of Honor*, 83 AM. J. INT'L L. 461 (1989).

79. *See generally* Martha Minow, *Interpreting Rights: An Essay for Robert Cover*, 96 YALE L.J. 1860, 1861 (1987) (arguing that law is "a communal language" and urging that it be interpreted in "social contexts in which norms can be generated and given meaning"); Howard Tolley, Jr., *Popular Sovereignty and International Law: ICJ Strategies for Human Rights Standard Setting*, 11 HUM. RTS. Q. 561 (1989).

80. As Professor Sohn noted,

> [O]n one hand, acceptance of the lowest possible common denominator would assure rapid ratification, but the documents would have no real effect; on the other hand, strict adherence to high ideals might lead states to refuse to ratify the documents, and the instruments would thus be of little value.

Louis B. Sohn, *The New International Law: Protection of the Rights of Individuals Rather Than States*, 32 AM. U. L. REV. 1, 39 (1982). *See generally* Ranee K.L. Panjabi, *Describing and Implementing Universal Human Rights*, 26 TEX. INT'L. L.J. 189 (1991)(reviewing JACK DONNELLY, UNIVERSAL HUMAN RIGHTS IN THEORY AND PRACTICE (1989)).

81. Optional Protocol to the International Covenant on Civil and Political Rights, *opened for signature* Dec. 19, 1966, 999 U.N.T.S. 302 (entered into force Mar. 23, 1976).

82. *E.g.*, Charter of the Organization of American States, Apr. 30, 1948, 2 U.S.T. 2394, 119 U.N.T.S. 3, *amended by* Feb. 27, 1967, 2 U.S.T. 607 *and* Dec. 5, 1985, 25 I.L.M. 529. *See generally* JACK DONNELLY, UNIVERSAL HUMAN RIGHTS IN THEORY AND PRACTICE (1989); Gerald L. Neuman, *Whose Constitution?*, 100 YALE L.J. 909 (1991)(discussing the evolution of the United States tradition of predicating constitutional rights on "membership" in the domestic polity and alternatives to that tradition in an international context).

83. *See, e.g.*, 1 HUMAN RIGHTS IN INTERNATIONAL LAW: LEGAL AND POLICY ISSUES 247-53 (Theodor Meron ed., 1984); Francesco Capotorti, *Human Rights: The Hard Road Towards Universality*, *in* THE STRUCTURE AND PROCESS OF INTERNATIONAL LAW: ESSAYS IN LEGAL PHILOSOPHY 986 (Ronald St. John MacDonald & Douglas M. Johnston eds., 1983). *See generally* Philip Alston, *The International Covenant on Economic, Social and Cultural Rights*, *in* MANUAL ON HUMAN RIGHTS REPORTING 39, U.N. Doc. HR/PUB/91/1, U.N. Sales No. E.91.XIV.1 (1991); Philip Alston, *The Purposes of Reporting*, *in* MANUAL ON HUMAN RIGHTS REPORTING, *supra*, at 13; John P. Humphrey, *The Implementation of International Human Rights Law*, 24 N.Y.L. SCH. L. REV. 31 (1978); Fausto Pocar & Cecil Bernard, *National Reports: Their Submission to Expert Bodies and Follow-Up*, *in* MANUAL ON HUMAN RIGHTS REPORTING, *supra*, at 25.

United States, for example, has been wary of being held to a more rigorous standard than less affluent countries.[84]

Because a state has no obligation to provide aid to another, bilateral aid can and has been conditioned on respect for certain human rights by the recipient state.[85] Similarly, because a state has no obligation to trade with another state, it can certainly refrain from doing so on human rights grounds. Once trade has been entered into, however, a state is not free to terminate the relationship because of subsequent human rights violations, however repugnant, of its trading partner. Violating the human rights of its own people does not justify countermeasures. Even humanitarian intervention—even after the Kurds[86]—remains highly sensitive.[87] Those denied their human rights depend primarily on the domestic legal system of the denying state for their vindication.[88] In some countries there is a "culture of compliance" and respect for the rule of law.[89] In states where human rights protections are most needed, however, there usually is not.[90]

---

84. *See* Thomas M. Franck, *Of Gnats and Camels: Is There a Double Standard at the United Nations?*, 78 AM. J. INT'L L. 811, 819-25 (1984).

85. *E.g.*, 22 U.S.C. § 262d(d) (Supp. I 1989)(United States to seek to channel economic assistance to governments respecting human rights); *see also Canada Announces Measures to Protest at Chinese Crackdown*, *supra* note 74 (Canada cancels 9.1 million dollars in aid to China following Tiananmen Square debacle).

86. *See* Tom J. Farer, *Human Rights and Foreign Policy: What the Kurds Learned (A Drama in One Act)*, 14 HUM. RTS. Q. 62 (1992).

87. As Professor O'Connell pointed out, however, "Distribution of humanitarian aid, even against the wishes of a government in effective control, is not unlawful intervention according to the International Court of Justice." O'Connell, *supra* note 30, at 906.

International law still rejects the use of force for humanitarian intervention by a state (except perhaps to save the lives of hostages, as the Israelis did at Entebbe). Henkin, *supra* note 8, at 151-52. "A different question is the permissibility of collective humanitarian intervention on the authority of the U.N. or of a regional body such as the [Organization of American States]." *Id.*

88. *See* Louis Henkin, *Rights: American and Human*, 79 COLUM. L. REV. 405 (1979); *accord* Bilder, *supra* note 76, at 13 ("Once again, the easiest and most effective way to implement human rights is through action within each nation's own legal system.").

89. Great Britain, for example, has a well-established tradition of deference to the rule of law. LOUIS HENKIN, THE RIGHTS OF MAN TODAY 51 (1978). *See generally* R.R. FENNESSEY, BURKE, PAINE, AND THE RIGHTS OF MAN 213-50 (1963)(describing schism in eighteenth century British liberal thought over the amount of deference due human rights considerations). However, even such states may have blind spots or lapses, as shown, for instance, by Britain's record with respect to Northern Ireland. *See* Ireland v. United Kingdom, 25 Eur. Ct. H.R. (ser. A)(1978)(Judgment of Jan. 18), *reprinted in* MARK W. JANIS & RICHARD S. KAY, EUROPEAN HUMAN RIGHTS LAW 117-32 (1990). *See generally Symposium on Human Rights in the U.S.*, 135 PROC. AM. PHILOSOPHICAL SOC'Y 1 (1991).

90. Jimmy Carter, *The Rule of Law and the State of Human Rights*, 4 HARV. HUM. RTS. J. 1, 3 (1991) (describing "too many countries [where] the final

Even if the domestic state endeavors to provide a remedy for a human rights violation, the victim is confronted with the second caveat: the human rights norm at issue is interpreted and enforced under domestic law.[91] Adherence to the Political Covenant represents a continuum of commitment to its principles. States differ, not only in the degree of deference they give international law,[92] but in how they interpret the covenant's provisions as applied in their respective domestic contexts.[93] While giving lip service to civil and political rights, for example, the Soviet form of these rights was virtually unrecognizable to a Western viewer.[94] Under the former Soviet constitution, civil and political rights were conceived of less as "negative rights," constraining the state, than as positive rights, granted (and determined) by the state itself. The people's "right to free association," for instance, consisted of an affirmative right to assemble in specific public buildings designated by the state.[95] While domestic construction of norms may accommodate legitimate concerns of cultural relativism,[96] human "dignity and worth" may become a

---

decisions are made by the government itself, depending on transient circumstances. . . . No higher law constrains the state."). *See generally* Robert F. Drinan & Teresa T. Kuo, *The 1991 Battle for Human Rights in China*, 14 HUM. RTS. Q. 21 (1992).

91. *See* STANDING COMM. ON WORLD ORDER UNDER LAW, AM. BAR ASS'N, INVOKING INTERNATIONAL HUMAN RIGHTS LAW IN DOMESTIC COURTS 16-18 (1985) (considering how international human rights law could "infuse" domestic standards); Richard B. Bilder, *Integrating International Human Rights Law into Domestic Law—U.S. Experience*, 4 HOUS. J. INT'L L. 1 (1981).

92. *See* Louis Henkin, *The Constitution and United States Sovereignty: A Century of* Chinese Exclusion *and Its Progeny*, 100 HARV. L. REV. 853 (1987) (treaties should not be subordinated to subsequent statutes). *But see* Peter Westen, *The Place of Foreign Treaties in the Courts of the United States: A Reply to Louis Henkin*, 101 HARV. L. REV. 511, 512 (1987)(lawful treaties are "lexically superior to statutes and . . . binding on the political branches of government").

93. As Henkin suggested, at the very least adherence represents an acknowledgement that we live in an "age of rights," an age in which rights have acquired universal cachet. HENKIN, *supra* note 53, at ix-x. *But see* Henry J. Steiner, *The Youth of Rights*, 104 HARV. L. REV. 917 (1990)(reviewing HENKIN, *supra* note 53). *See generally* John H. Jackson, *Status of Treaties in Domestic Legal Systems: A Policy Analysis*, 86 AM. J. INT'L L. 310, 313 (1992)(discussing "status of treaties in national legal systems, that is, the question of 'direct application'; and the hierarchical status in national legal systems when directly applied treaty norms clash with other norms of the same system").

94. The Soviet Union was a party to the Civil Covenant. HENKIN ET AL., BASIC DOCUMENTS SUPPLEMENT TO INTERNATIONAL LAW: CASES AND MATERIALS 388 (2d ed. 1987). For a scholarly comparison of the Soviet and American approaches to domestic implementation of human rights norms, see Damrosch, *supra* note 6.

95. Konstitusiia SSSR [Constitution] art. 50 (U.S.S.R.)(1977). *See* Damrosch, *supra* note 6, at 2330 & n.77 ("International human rights law resists the tendency of Soviet constitutional law to place the interests of the state above the rights of individuals.")

96. *See generally* RELATIVISM: INTERPRETATION AND CONFRONTATION, *supra* note 14; Donnelly, *supra* note 14; Renteln, *supra* note 14.

variable and even indeterminable standard, depending on one's location and the current political situation in that state.

Implementation of economic rights is even more problematic. The major mechanism for assuring compliance, aside from domestic law, is the preparation and submission of self-monitoring reports to the United Nations Committee on Economic, Social and Cultural Rights (Committee).[97] The Committee meets with state representatives after its review of the reports. During this meeting, which is open to the public, the Committee typically asks for further information or clarification[98] and concludes with comments[99] intended to enable the state to better achieve its own objectives.[100] A similar self-

---

97. As Professor Alston noted, "The principal obligation of States parties to the [Economic] Covenant is to implement its provisions at the national level. The obligation to report to an international body . . . is essentially a means of promoting the implementation of that obligation." Alston, *supra* note 83, at 39. The Committee also uses the reporting process "to demonstrate a consistency of approach from one report to another." *Id.* at 40. This procedure has not always been effective. *See* Forsythe, *supra* note 69, at 541 (East Europeans resisted Committee review prior to collapse of Soviet bloc); *accord* REBECCA M.M. WALLACE, INTERNATIONAL LAW 189-90 (1986). *See generally* Alston, *supra* note 83, at 13-16, 39-77; Humphrey, *supra* note 83, at 37-38.

Article 16 of the Economic Covenant requires the parties to submit "reports on the measures which they have adopted and the progress made" to the Secretary-General of the United Nations. Economic Covenant, *supra* note 55, art. 16, paras. 1-2. The Secretary-General originally transmitted copies of the reports to the United Nations Economic and Social Council, but they are now submitted directly to the Committee. For a full account of the reasons for the change, and its consequences, see Philip Alston & Bruno Simma, *Second Session of the UN Committee on Economic, Social and Cultural Rights*, 82 AM. J. INT'L L. 603 (1988); *see also* MOWER, *supra* note 69, at 31-46. For present purposes, it is sufficient to note that the formation of an independent monitoring organ represented both an acknowledgement of the inadequacy of the original system and a renewed commitment to economic rights on the part of the United Nations.

98. Countries are accordingly encouraged to send knowledgeable experts to these meetings. Interview with Alexandre Tikhonov, Secretary to the Committee, United Nations Centre for Human Rights, Palais des Nations, in Geneva, Switz. (June 11, 1991).

99. Pocar & Bernard, *supra* note 83, at 25, 26.

100. The Committee indicates when a report, or the activity reported, fails to satisfy ICESCR. Interview with Alexandre Tikhonov, *supra* note 98; *see also* Alfred de Zayas, *The Potential for the United States Joining the Covenant Family*, 20 GA. J. INT'L & COMP. L. 299, 304 (1990)("[D]iscussions have been serious, well-focused, and non-political. . . . [T]he Committee has encouraged but not pressured states parties."). For an example of relatively vigorous questioning, see *Report on the Third Session*, U.N. Committee on Economic, Social and Cultural Rights, 3d Sess., at 34, U.N. Doc. E/1989/22, E/C.12/1989/5 (France questioned about right to housing).

The Committee also prepares "general comments," which are not limited to specific countries. "The Committee endeavors, through its general comments, to make the experience gained so far through the examination of these reports available for the benefit of all States parties in order to assist and promote their further

monitoring regime has been proposed in connection with international environmental rights.[101]

III. ENVIRONMENTAL RIGHTS

A. Impetus for Acceptance

Environmental rights are rooted in both our real fears of collective annihilation and the growing international recognition of the dignity and worth of the individual human being.[102] Like peace, environmental rights have emerged from the common realization that our survival as a species, as well as the survival of the other species with whom we share this planet, requires their recognition and rigorous enforcement.[103] Indeed, it may be argued that the threat of annihi-

---

implementation of the [Economic] Covenant." *Id.* Annex III, at 87.

While this illustrates the deference accorded a sovereign state under international enforcement procedures, it should be kept in mind: "As is not the case with civil-political rights, . . . another state can help give effect to some economic-social rights . . . without forcible intervention, merely by financial aid to the local government." HENKIN, *supra* note 53, at 45. *But see* Lloyd N. Cutler, *The Internationalization of Human Rights*, 1990 *U. Ill. L. Rev.* 575, 588 ("economic rights are especially unsuitable for international protection by one state against another or by the international community as a whole").

101. *See* Oscar Schachter, *The Emergence of International Environmental Law*, 44 J. INT'L AFF. 457 (1991). "[I]n what many said would be the true significance of [the Earth Charter], machinery would be set up to constantly assess the danger of climate change and to take further action, if necessary." William K. Stevens, *43 Lands Adopt Treaty to Cut Emissions of Gases*, N.Y. TIMES, May 10, 1992, § 1, at 14.

102. For a cogent introduction, see Richard B. Bilder, *The Settlement of Disputes in the Field of the International Law of the Environment*, 144 RECUEIL DES COURS 139, 145-50 (1975). For astute assessments of the existing regimes, see Palmer, *supra* note 16; Catherine Tinker, *Environmental Planet Management by the United Nations: An Idea Whose Time Has Not Yet Come?*, 22 N.Y.U. J. INT'L L. & POL. 793 (1990). For a comprehensive overview, see Developments in the Law, *International Environmental Law*, 104 HARV. L. REV. 1484 (1991). For an authoritative analysis, see Schachter, *supra* note 101. *See generally* FACTSHEET: THE UNITED NATIONS AND THE GLOBAL ENVIRONMENT (United Nations Ass'n of the United States of America, New York, N.Y.).

About 140 multilateral treaties on environmental issues had been concluded as of 1990. Schachter, *supra* note 101, at 470. For a succinct description of some of the major international treaties, resolutions, and declarations, see A.O. Adede, *A Profile of Legal Instruments for International Responses to Problems of Environment and Development*, 21 ENVTL. POL'Y & L. 224 (1991).

103. AL GORE, EARTH IN THE BALANCE: ECOLOGY AND THE HUMAN SPIRIT (1991)(civilization cannot survive unless saving the environment becomes its organizing principle); Catherine Tinker, *Environmental Damage and the United Nations Security Council: Towards a Broad Definition of Threats to International Peace and Security and the Need for Collective Security*, 59 TENN. L. REV. 787 (1992); *cf.* William K. Stevens, *Lessons of Rio: A New Prominence and an Effective Blandness*, N.Y. TIMES, June 14, 1992, § 1, at 10. (noting "new-found prominence of the environment as an international issue, bidding to rank with economics and national security").

lation from continuing violations of environmental rights is less of a risk and more of a certainty than nuclear war ever was.

It may also be more urgent.[104] A major part of the nuclear horror was the potential for catastrophe without time to prepare, to warn, or to avoid. Advocates of disarmament emphasized the risk of instant holocaust, even instant holocaust by mistake.[105] The Cuban missile crisis was probably the closest we came—surely close enough, but in fact it was avoided.[106]

Environmental apocalypse is qualitatively distinguishable. Even if absolute catastrophe may be averted,[107] even if policies and practices which lead to it are abandoned, we have a long way to go to recovery.[108] We must cope with massive clean-ups and remote, often unforeseen, consequences. Reclamation may not only be dauntingly complex and expensive, but impossible.[109] While environmental destruction may be even more pressing than the threat of war, it may well be less susceptible to diplomatic resolution or deterrence. Private actors, for example, are likely to be less easily controlled by the state than its own personnel.

Equally important, the international commitment to environmental rights is compelled by the same considerations of "human worth

---

104. The threat posed by humans to the global environment may well be the major danger in the post-Cold War world. *See* GORE, *supra* note 103, at 34-35. The Bush Administration's lack of a coherent environmental policy "raise[s] questions about the United States role in a world in which national security may be as affected by global environmental threats as by military ones." Keith Schneider, *Environmental Policy: It's a Jungle in There*, N.Y. TIMES, June 7, 1992, § 4, at 1.

105. *See, e.g.*, EUGENE BURDICK, FAIL-SAFE (1962). As Anthony D'Amato remarked, "Acts of cosmic stupidity are always possible . . . ." Anthony D'Amato, *Do We Owe a Duty to Future Generations to Preserve the Global Environment?*, 84 AM. J. INT'L L. 190, 190 n.6 (1990).

106. *See generally* ABRAM CHAYES, THE CUBAN MISSILE CRISIS (1974); Brunson MacChesney, *Some Comments on the "Quarantine" of Cuba*, 57 AM. J. INT'L L. 592 (1963); Quincy Wright, *The Cuban Quarantine*, 57 AM. J. INT'L L. 546 (1963).

107. It has been suggested, for example, that space may serve as a "safety-valve"—that within "perhaps two human lifetimes, it will be possible to move most polluting industries off the Earth and into space. And the industries that remain can be made far less polluting through the use of clean, inexpensive energy derived from space." OUTER SPACE AND THE GLOBAL ENVIRONMENT: A NSS POSITION PAPER 4-5 (Nat'l Space Soc'y, Washington, D.C.). For a concise overview of environmental concerns in outer space, see GLENN H. REYNOLDS & ROBERT P. MERGES, OUTER SPACE: PROBLEMS OF LAW AND POLICY 195-98 (1989).

As Professor Schachter pointed out, extension of the concept of environmental harm to outer space presents as yet unresolved issues of policy. Schachter, *supra* note 101, at 465-66.

108. Douglas R. Weiner, *Chernobyl Isn't the Whole Story*, N.Y. TIMES, June 7, 1992, § 7, at 14 (describing "full recovery" for the Soviet Union as an "unimaginably expensive prospect").

109. Schachter, *supra* note 101, at 472-73 (discussing the irreversibility of global warming).

and dignity" that mandate compliance with other human rights norms. As the environment deteriorates, there is a growing acknowledgement that health as well as the use and enjoyment of natural resources are as crucial to the realization of human dignity as first and second generation human rights.[110] Environmental rights, moreover, meet the United Nations General Assembly guidelines for states and United Nations bodies articulating new rights. These rights are "consistent with . . . existing . . . international human rights . . . of fundamental character and derive from the inherent dignity and worth of the human person . . . sufficiently precise . . . provide . . . realistic and effective implementation . . . [and] . . . [a]ttract broad international support."[111] Indeed, Stephen Marks has described environmental rights as

> the most "classical" case of a set of claims which have been given holistic formulation in terms of human rights. All the features of a right of the new generation are there: elaboration of a specialized body of law, an easily identifiable international legislative process, incorporation of the right as human right within municipal systems, and need for concerted efforts of all social actors.[112]

B. Concessions to State Sovereignty[113]

Unlike the leeway given states in connection with intrastate use of force and the dearth of interstate enforcement of human rights,

---

110. For an early appreciation, see Charles Maechling, *The Emergent Right to a Decent Environment*, 1 HUM. RTS. 59 (1970). The first important intergovernmental meeting to address the need for an international response to environmental degradation was held in Stockholm in 1972. Declaration of the U.N. Conference on the Human Environment in Stockholm, June 16, 1972, 11 I.L.M. 1416 (1972), U.N. Doc. A/CONF. 48/14/Rev. 1. (1974) [hereinafter Stockholm Declaration]; *see also* Louis B. Sohn, *The Stockholm Declaration on the Human Environment*, 14 HARV. J. INT'L L. 423 (1973). *See generally* W. PAUL GORMLEY, HUMAN RIGHTS AND ENVIRONMENT: THE NEED FOR INTERNATIONAL CO-OPERATION 121-45 (1976) (describing United Nations "experiments" to formulate a functional program).

This refers to future as well as present health, use, and enjoyment. *See* EDITH BROWN WEISS, IN FAIRNESS TO FUTURE GENERATIONS: INTERNATIONAL LAW, COMMON PATRIMONY AND INTERGENERATIONAL EQUITY 1-3 (1989); RESPONSIBILITIES TO FUTURE GENERATIONS (E. Partridge ed., 1981); Agora, *What Obligation Does Our Generation Owe to the Next? An Approach to Global Environmental Responsibility*, 84 AM. J. INT'L L. 190 (1990). *See generally* THE FUTURE OF THE INTERNATIONAL LAW OF THE ENVIRONMENT (René-Jean Dupuy ed., 1984).

111. *Setting International Standards in the Field of Human Rights*, G.A. Res. 120, U.N. GAOR, 41st Sess., Supp. No. 53, at 178, U.N. Doc. A/41/53 (1987).

112. Marks, *supra* note 12, at 442-43.

113. For a cogent description of approaches to the "sovereignty problem" in this context, i.e., obtaining the assent and assuring the compliance of sovereign states, see Developments in the Law, *supra* note 102, at 1552-66.

there are no accepted structural exceptions to international environmental rights.[114] The states of the Southern Hemisphere argue that allowances must be made in order to permit them to develop and provide their people with a standard of living more like that of the already industrialized North.[115] The well-established law against transboundary pollution,[116] in conjunction with the already alarming contamination of the global environment, has made the Northern states unreceptive to these arguments, especially in view of their own decreasing reliance on heavy industry and their comfortable economic hegemony.[117]

---

114. Sovereignty concerns are typically framed in terms of "States' sovereign rights over their own natural resources." Stockholm Declaration, *supra* note 110, principle 21; *accord Developing Countries and International Environmental Law*, 21 ENVTL. POL'Y & L. 213 (1991)(formulation adopted at symposium sponsored by Chinese Government). The implications of sovereignty in this context may have grave consequences. *See, e.g.*, Daniel B. Magraw, *Transboundary Harm: The International Law Commission's Study of "International Liability,"* 80 AM. J. INT'L L. 305, 325 (1986)(importation of hazardous waste as a sovereign prerogative of a developing state); Jeffery D. Williams, Comment, *Trashing Developing Nations: The Global Hazardous Waste Trade*, 39 BUFF. L. REV. 275 (1991); Symposium, *The Bhopal Tragedy: Social and Legal Issues*, 20 TEX. INT'L L.J. 269 (1985); *cf.* OSCAR SCHACHTER, INTERNATIONAL LAW IN THEORY AND PRACTICE 383 (1991)("[L]inkage between international and domestic law is crucial for environmental protection.").

115. *See, e.g.*, Ved P. Nanda, *International Environmental Protection and Developing Countries' Interests: The Role of International Law*, 26 TEX. INT'L L.J. 497 (1991)(reviewing developing countries' perspectives on the export of hazardous wastes and pesticides and protection of the ozone layer); Symposium, *International Development Agencies (IDAs), Human Rights and Environmental Considerations*, 17 DENV. J. INT'L L. & POL'Y 29 (1988).

Even on the national level, consensus is often difficult to achieve. *See, e.g.*, David M. Driesen, *The Congressional Role in International Environmental Law and Its Implications for Statutory Interpretation*, 19 B.C. ENVTL. AFF. L. REV. 287 (1991).

116. Trail Smeltor Case (U.S. v. Can.), III R.I.A.A. 1905 (U.N. Arbitral Trib. 1949); Stockholm Declaration, *supra* note 110, principle 21; *see also* Corfu Channel Case (U.K. v. Alb.), 1949 I.C.J. 4, 22 (April 9) (holding that every state has an obligation "not to allow knowingly its territory to be used for acts contrary to the rights of other States"—Albania's failure to warn a British ship of mines in the Corfu Channel was accordingly a violation of international law). *See generally* Magraw, *supra* note 114.

117. I am not suggesting that the "Northern states" have agreed upon a consistent approach. *See* Steven Keeva, *Environmental Law Takes Root*, 78 A.B.A. J., May 1992, at 52, 54. For a description of the process through which consensus is sought among the members of the European Community, see Michael S. Feeley & Peter M. Gilhuly, *Green Law-Making: A Primer on the European Community's Environmental Legislative Process*, 24 VAND. J. TRANSNAT'L L. 653 (1991). *See generally*, David A. Westbrook, *Environmental Policy in the European Community: Observations on the European Environment Agency*, 15 HARV. ENVTL. L. REV. 257 (1991); G. Nelson Smith, III, *A Comparative Analysis of European and American Environmental Laws: Their Effects on International Blue Chip Corporate Mergers*

The draft Earth Charter,[118] prepared for the United Nations Conference on Environment and Development (UNCED),[119] held in June 1992, reflects some important rhetorical compromises.[120] Principle 3, for example, provides that "[T]he right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations."[121] Principle 4, similarly, concedes both the need for "sustainable development" and the need for "environmental protection" within the framework of any such development: "In order to achieve sustainable development, environmental protection shall constitute an integral part of the development process and cannot be considered in isolation from it."[122]

As Professor Edith Brown Weiss recently observed,[123] the major issue for UNCED is how these interests may be addressed in the context of an effective international regime.[124] While there may be

---

and Acquisitions, 14 HASTINGS INT'L & COMP. L. REV. 573 (1991).

Neither the "North" nor the "South" is monolithic. William K. Stevens, *Rio: A Start on Managing What's Left of This Place*, N.Y. TIMES, May 31, 1992, § 4, at 1. The United States, like some "Southern" states, rejected measures that it feared would adversely affect economic growth. For example, the Bush administration feared the biodiversity treaty would damage the biotechnology industry. Steven Greenhouse, *Ecology, the Economy and Bush*, N.Y. TIMES, June 14, 1992, § 4, at 1.

118. Declaration of the U.N. Conference on Environment and Development in Rio de Janeiro, April 4, 1992 (draft) [hereinafter Draft Rio Declaration] *reprinted in Draft of Environmental Rules: "Global Partnership,"* N.Y. TIMES, April 5, 1992, § 1, at 10.

119. *See* G.A. Res. 228, U.N. GAOR, 44th Sess., Supp. No. 49, at 151, U.N. Doc. A/44/49 (1990). Exhaustive planning for UNCED included three meetings of the Preparatory Committee (PrepCom I, II, and III). *See PrepCom: Third Meeting*, 21 ENVTL. L. & POL'Y 186 (1991). During these meetings the broad structure for UNCED was endorsed. This structure consists of three elements: a statement of principles (the Earth Charter), an agenda for action in the 21st century (Agenda 21), and global treaties, including treaties on climate change and biodiversity. *Id.*

For a useful overview of UNCED in nontechnical terms, see Bruce Babbitt, *The World After Rio*, WORLD MONITOR, June 1992, at 28.

120. *But see* Martti Koskenniemi, *The Future of Statehood*, 32 HARV. INT'L L.J. 397, 403 (1991).

> The official ideology of [UNCED] compels diplomats to speak of environmental and developmental goals as if there were no essential conflict between them, by defining one in terms of the other. Poverty is pollution; environmental quality is an aspect of the standard of living. Such harmony is soon dispelled when concrete action is debated.

Id.

121. Draft Rio Declaration, *supra* note 118, principle 3.

122. *Id.* principle 4.

123. Edith Brown Weiss, Remarks at the American Society of International Law Annual Meeting (January 3, 1992).

124. The normative authority of UNCED is an open question. *See* THOMAS M. FRANCK, THE POWER OF LEGITIMACY AMONG NATIONS 16 (1990) (describing how "a rule or rule-making institution . . . itself exerts a pull towards compliance on

some room for accommodating different local conditions, needs, and standards, it is doubtful that variation on the scale tolerated under other human rights treaties could be accepted in the context of environmental rights.[125] The argument can be made that such variation should not be tolerated in any context—that this amounts to the derogation of purportedly agreed-upon norms and the ultimate subversion of the protective regime. It has been suggested that the Women's Convention (CEDAW),[126] for example, is so riddled with reservations and understandings[127] that the international regime itself is debased.[128] The extent of such debasement is unclear, however, in

---

those addressed normatively"); Richard L. Williamson, Jr., *Building the International Environmental Regime: A Status Report*, 21 U. MIAMI INTER-AM. L. REV. 679 (1990)(giving an overview of key international environmental problems and assessing the international response to those problems); *see also* Roberta Dohse, Comment, *Global Air Pollution and the Greenhouse Effect: Can International Legal Structures Meet the Challenge?*, 13 HOUS. J. INT'L L. 179 (1990)(global air pollution and the greenhouse effect); Ved P. Nanda, *Stratospheric Ozone Depletion: A Challenge for International Environmental Law and Policy*, 10 MICH. J. INT'L. L. 482 (1989)(analyzing the possibilities and limitations of the international regime for the protection of the ozone layer). For a detailed account of the first global-international (as opposed to regional-interstate) treaty for environmental protection, see VEIT KOESTER, THE RAMSAR CONVENTION ON THE CONSERVATION OF WETLANDS (1989). *See generally* Myron L. Scott, *Two Models for Environmental Cooperation*, 22 ENVTL. L. 349 (1992)(reviewing KOESTER, *supra*, and PETER M. HAAS, SAVING THE MEDITERRANEAN: THE POLITICS OF INTERNATIONAL ENVIRONMENTAL COOPERATION (1990)).

125. *See, e.g.*, Michael B. Saunders, Comment, *Valuation and International Regulation of Forest Ecosystems: Prospects for a Global Forest Agreement*, 66 WASH. L. REV. 871, 891 (1991)("Previous agreements designed to protect global resources located within national borders have proven to be of limited effectiveness. States perceive conflicts between economic interests and conservation and fail to undertake measures that they believe are incompatible with national sovereignty.").

126. Convention on the Elimination of All Forms of Discrimination Against Women, Dec. 18, 1979, 1249 U.N.T.S. 13 (entered into force Sept. 3, 1981).

127. *See* Vienna Convention on the Law of Treaties, *supra* note 76, arts. 19, 31 (formulation of reservations and interpretation of treaties). *See generally* Belilos Case, 132 Eur. Ct. H.R. (ser. A) at 21-24 (1988); D.W. Bowett, *Reservations to Non-Restricted Multilateral Treaties*, 48 BRIT. Y.B. INT'L. L. 67 (1976-1977).

128. *See* Belinda Clark, *The Vienna Convention Reservations Regime and the Convention on Discrimination Against Women*, 85 AM. J. INT'L L. 281 (1991); Rebecca J. Cook, *Reservations to the Convention on the Elimination of All Forms of Discrimination Against Women*, 30 VA. J. INT'L. L. 643 (1990). States' often haphazard observance of CEDAW exacerbates this debasement. International consensus on gender discrimination is notably problematic. For a comprehensive and perceptive discussion, see Hilary Charlesworth et al., *Feminist Approaches to International Law*, 85 AM. J. INT'L L. 613 (1991). *But see* Arthur Rovine & Jack Goldklang, *Defense of Declarations, Reservations, and Understandings*, *in* U.S. RATIFICATION OF THE HUMAN RIGHTS TREATIES: WITH OR WITHOUT RESERVATIONS? 54 (Richard B. Lillich ed., 1981); Reservations to the Convention on Prevention and Punishment of the Crime of Genocide, 1951 I.C.J. 15 (May 28). *See generally* Rebecca J. Cook, *International Human Rights Law Concerning Women: Case Notes*

part because its perception is so contingent upon cultural factors. As a corollary, there is no agreed-upon method of precisely measuring derogation.

Environmental degradation, in contrast, is often demonstrably uncontainable,[129] and the spread of pollution may be ascertained with relative precision.[130] For this same reason, interstate tolerance of violations, deplored but characteristic of first and second generation human rights, is not feasible in the environmental context, at least not to the extent hitherto condoned with respect to other human rights. As Sir Geoffrey Palmer warned, "The stakes are so high that slippage in meeting the standards will be intolerable. The actions of one nation could render nugatory the actions of all the others to preserve the global environment."[131] At the same time, however, states for the most part remain as reluctant to submit to the judgment of other states as they are when human rights violations are claimed.[132]

Environmental rights are predicated on both the concern for our collective survival underlying the commitment to international peace

---

and Comments, 23 VAND. J. TRANSNAT'L. L. 779 (1990). For a cogent analysis of this problem in the political rights context, see Oscar Schachter, *The Obligation of the Parties to Give Effect to the Covenant on Civil and Political Rights*, 73 AM. J. INT'L L. 462 (1979).

129. It has become "axiomatic that these problems transcend the capacity of any nation to handle or avoid." Bilder, *supra* note 102, at 146 (quoting Maurice F. Strong, *One Year After Stockholm*, 51 FOREIGN AFF. 690, 697 (1973)); *accord*, FACTSHEET: THE UNITED NATIONS AND THE GLOBAL ENVIRONMENT, *supra* note 102, at 4 (quoting Carl Sagan, Remarks at the Global Forum on Environment and Development in Moscow (Jan. 15-19, 1990)) ("Intrinsically, [these assaults on the environment] are transnational, transgenerational, and trans-ideological. So are all conceivable solutions."). *See generally* Robert A. Kaplan, *Into the Abyss: International Regulation of Subseabed Nuclear Waste Disposal*, 139 U. PA. L. REV. 769 (1991).

This does not mean, of course, that there are always "significant or substantial" harmful effects, which will necessarily cross national boundaries. Schachter, *supra* note 101, at 463-65. Thus, not all "degradation" would be cognizable under existing international regimes. *Id.*

130. At the same time, the significant difficulties confronting environmentalists should not be underestimated. Monitoring is not simple, and in too many substantive (as well as geographic) areas it is not being done at all. Palmer, *supra* note 14, at 263. What are the standards and what is the process for determining them? What are acceptable deviations? Who has the responsibility and who has the authority to decide these questions? *See, e.g.*, Stevens, *supra* note 101 (describing unsuccessful efforts to "establish clear targets and timetables on emissions"). These are not abstract problems, but ever-present dilemmas in negotiation. For a discussion of possible legal approaches to the problem of setting specific standards, see Schachter, *supra* note 101, at 467.

131. Palmer, *supra* note 16, at 282.

132. *See, e.g.*, *supra* notes 14, 33, 75. The idea of international "green policing," for example, which is central to the plan of the European Environment Agency, has been criticized for interfering with member state sovereignty. Westbrook, *supra* note 117, at 263-64.

and the concern for individual dignity and worth underlying international human rights. Because environmental rights require compatible intrastate and interstate standards and compliance mechanisms, however, they place unprecedented demands on the state system.

CONCLUSION

In his brilliant novel, *The Remains of the Day*,[133] Kazuo Ishiguro describes an automobile trip taken by an aging butler through the English countryside.[134] Inspired by the natural beauty around him, the butler finds himself reviewing his life, particularly the years between the world wars. As he wanders through the tranquil landscape he thinks about the failure of the world leaders to avert the second world war and his own unquestioning support of what he now realizes were their misguided efforts. At the end of his journey, he sits at dusk by the ocean with tears streaming down his face.[135] A stranger tries to comfort him: "The evening's the best part of the day. You've done your day's work. Now you can put your feet up and enjoy it."[136]

But Ishiguro's hero has spent a lifetime serving in what he refers to without irony as one of the "big houses,"[137] passively accepting rules, hierarchies, and boundaries that not only kept him from exploring the natural world but from meaningful human contact as well. He has had a belated glimpse of a harmonious natural world and his place in it, a clear view of what he has already lost, but he seems more likely to suffer than to learn from his vision.

We, too, have glimpsed an integrated vision of a harmonious natural world and our place in it.[138] We have to ask whether a system of sovereign states can support such a vision.[139] The state system has

---

133. KAZUO ISHIGURO, THE REMAINS OF THE DAY (1989).

134. *Id.*

135. *Id.* at 240-45.

136. *Id.* at 243-44.

137. *Id.* at 241.

138. I am not suggesting that we have all had such a vision personally. But even if we have not—and even if we have not read Lester R. Brown, Rachel Carson, Annie Dillard, Christopher Stone, Edith Brown Weiss, or the Club of Rome's publications—we have followed the Earth Summit in the news, seen the Sierra Club calendar, "saved the whales," bought "dolphin-safe" tuna, recycled newspapers and aluminum cans, participated in Earth Day, or sat through FERN GULLY (Twentieth Century Fox 1992) or *Captain Planet* (ABC television broadcast, Saturday mornings). Environmental consciousness has become part of our zeitgeist.

139. [W]e must reassess our unquestioned respect for national sovereignty and our faith in the capacity of the nation-state to respond fully to the challenges we face. There are two areas . . . where I think this reality strikes hardest. One is human rights and the second is the protection of the environment.

Mondale, *supra* note 17, at 450.

shown that it can recognize and endorse universal values, but not without major structural concessions to state sovereignty.

Even if such concessions are acceptable in the contexts of peace and human rights, they are unendurable in the context of environmental rights. We see, in our dead rivers[140] and dying forests,[141] in the encroaching deserts of sub-Saharan Africa,[142] in the putrid stench of East European cities,[143] what we have lost.[144] We see the limitations that inhere in a sovereign system. While these limitations permit a pinched success in the peace and human rights regimes, they tolerate and perpetuate normative conflicts between intrastate and interstate regimes fatal to any meaningful conception of environmental rights. Are we capable of the transformative act of imagination necessary to articulate that conception?[145] Do we have the political will to realize it?[146] Unless both questions can be answered affirmatively—

---

140. *See, e.g.*, Ray Moseley, *E. Germans Fear Ecological Crisis*, CHI. TRIB., Feb. 4, 1990, § 1, at 23. (noting that the Elbe, "Europe's most polluted river," carries about 27 tons of mercury a year); Matt Neufeld, *Pols on the River Push for Cleaner Anacostia Tide*, WASH. TIMES, Aug. 28, 1990, at B4 (describing the "most polluted river on Chesapeake Bay").

141. *See, e.g.*, Timothy Egan, *Satellite View: Forest Damage, North and South*, N.Y. TIMES, June 14, 1992, § 4, at 6 (photos of forests from space show nearly 90 percent of the original Northwest forest is gone); Saunders, *supra* note 125.

142. Bruce Finley, *Desertification: Africans Losing Battle Against the Sahara*, SAN FRANCISCO CHRON., Apr. 20, 1992, at A10; Mark Huband, *Desert Creeps Up on Northern Nigeria*, THE GAZETTE (Montreal), June 2, 1992, at A14.

143. *See, e.g.*, MURRAY FESHBACH & ALFRED FRIENDLY, JR., ECOCIDE IN THE U.S.S.R.: HEALTH AND NATURE UNDER SIEGE (1992); Weiner, *supra* note 108.

144. *See generally* CLIVE PONTING, A GREEN HISTORY OF THE WORLD: THE ENVIRONMENT AND THE COLLAPSE OF GREAT CIVILIZATIONS (1992).

145. "[M]ethods and techniques now available to fashion new instruments of international law to cope with global environmental problems cannot meet [the] challenge." Palmer, *supra* note 16, at 264; *accord* Hermann Scheer, *Earth Summit in Rio: Will It Do More Harm than Good?*, THE NATION, Apr. 20, 1992, at 522, 523 ("At best [Rio will produce] nonmandatory, ineffective guidelines. The international political system is not capable of more—not now and not in the near future.").

146. Governor Babbitt correctly predicted a "three-part North-South bargain," consisting of a Northern commitment to stabilize, and then reduce, carbon dioxide emissions, and in exchange for some kind of Northern support for sustainable development in the South, Southern acceptance of a biodiversity treaty to stop the destruction of the rain forests and the extinction of plants and animals that live there. Babbitt, *supra* note 119, at 30.

Perhaps if we can rethink the limitations of the state system as opportunities for nonstate participation, including participation by international organizations and even individuals, we can begin to create a sustainable future. *See* Janis, *supra* note 10, at 363 (noting the "obvious importance of non-state actors in international politics"); *see also* Developments in the Law, *supra* note 102, at 1600-04 (urging broadened participation in the decision-making process). For a specific and creative example, see David A. Wirth, *Legitimacy, Accountability, and Partnership: A Model*

and soon—we, too, are likely to spend "the remains of the day" in futile regret.[147]

---

for Advocacy on Third World Environmental Issues, 100 YALE L.J. 2645 (1991). For a description of the ways in which scientists and conservationists may play a more significant role than states in shaping the law, see PETER M. HAAS, SAVING THE MEDITERRANEAN: THE POLITICS OF INTERNATIONAL ENVIRONMENTAL COOPERATION (1990), and Thorme, *supra* note 15, at 305-08 (Sierra Club Legal Defense Fund, along with Friends of the Earth, brought two environmental cases before the United Nations Sub-Commission on the Prevention of Discrimination and Protection of Minorities). Finally, of course, there is the Voluntary Human Extinction Movement. Under the slogan, "The Answer to All Our Problems," the movement points out that "the extinction of *Homo Sapiens* would mean survival for millions, if not billions, of other Earth-dwelling species." Theodore Roszak, *Green Guilt and Ecological Overload*, N.Y. TIMES, June 9, 1992, at A27.

See, e.g., DONELLA H. MEADOWS ET AL., BEYOND THE LIMITS: CONFRONTING GLOBAL COLLAPSE, ENVISIONING A SUSTAINABLE FUTURE (1992)(computer model-based argument for sustainability); Lester R. Brown, *Environment: World Watcher's Warning*, 5 WORLD MONITOR 18, 20 (May 1992). *But see* Developments in the Law, *supra* note 102, at 1639 (concluding that "the ultimate goal . . . must remain the development and strengthening of each state's own regulatory regime"); *accord* Melissa Thorme, *Local to Global: Citizen's Legal Rights and Remedies Relating to Toxic Waste Dumps*, 5 TUL. ENVTL. L.J. 101, 148-51 (1991)(noting that domestic options—on the local, state, and federal levels—are more promising options for dealing with toxic waste dumps than the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal, *opened for signature* March 22, 1989, S. TREATY DOC. No. 5, 102d Cong., 2d Sess., 28 I.L.M. 649 (entered into force May 5, 1992)); *cf.* Williamson, *supra* note 124, at 744-50 (making a persuasive case for addressing environmental problems on the "proper level," i.e., global, regional, bilateral, or national). *See generally* Richard B. Bilder, *The Role of Unilateral State Action in Preventing International Environmental Injury*, 14 VAND. J. TRANSNAT'L L. 51 (1981).

147. As Albert Schweitzer predicted, "Man has lost the capacity to foresee and forestall. He will end by destroying the earth." RACHEL CARSON, SILENT SPRING v (1962)(quoting Albert Schweitzer in dedication). Maybe not. For heartening descriptions of "paradigmatic success stories," see Scott, *supra* note 124.

Prerequisite to Peace: An International Environmental Ethos

ALAN L. BUTTON*

INTRODUCTION

Is pollution of the seas a problem? Is the decline of the polar bear a problem? Is desertification a problem? How do we decide whether these are problems? How important a problem is pollution of the seas? The decline of the polar bear? Desertification? Where do we look to decide how important these problems may be? Given their importance, how should they be addressed? Why should pollution of the seas or the decline of the polar bear or desertification be addressed in one way and not in another?

At various times people have come together to address each of these areas. In doing so, they have characterized all three—pollution of the seas, the decline of the polar bear, and desertification—as problems. They have made some attempt to resolve these problems, and their efforts have been peaceful. Indeed, one of their broader objectives, implicitly, has been peace.

These areas of international environmental interest have been addressed in at least three different ways. In the Declaration of the United Nations Conference on the Human Environment, more commonly known as the Stockholm Declaration, delegates from scores of countries around the world agreed that pollution of the seas is a problem and declared that it should be prevented.[1] In the Agreement on the Conservation of Polar Bears, five countries in possession of Arctic territories agreed that the decline of the polar bear is a problem and that trade in polar bears and their parts would be restricted.[2] In House Concurrent Resolution 248, the United States Congress agreed

---

* Associate Professor of Law, Norman Adrian Wiggins School of Law, Campbell University. I.L.M. 1989, University of Cambridge; J.D. 1981, Washington and Lee University; A.B. 1976, Cornell University. The author is especially grateful to Campbell law students James Edward Midkiff (1992), Wayne Authur Roper (1993), Charles E. Simpson, Jr. (1993) and Cynthia Balogh Thomas (1993) for their research assistance.

1. Declaration of the United Nations Conference on the Human Environment, princ. 7, 11 ILM 1416 (1972)[hereinafter Stockholm Declaration]. *See infra* text accompanying notes 12-68.

2. Agreement on the Conservation of Polar Bears, Nov. 15, 1973, 27 U.S.T. 3918. *See infra* text accompanying notes 97-98.

that desertification can lead to political instability and urged that United States foreign assistance be used to address the consequences of desertification.[3]

The people involved in each of these endeavors have agreed that they share a common concern and that a particular mechanism is appropriate for the resolution of the problems involved. Agreement between the interested parties is a basic element of each mechanism. Unfortunately perhaps, given the relative vagueness of the phraseology used in the agreements, as well as the general lack of direct enforcement power of international law, there is much room for maneuvering on the part of the obligees.

For example, House Concurrent Resolution 248, titled "Sense of Congress Regarding Linkage Between the Environment and National Security," affirms the connection between the environment and international peace.[4] Congress has urged the Secretary of State and the Administrator of the Agency for International Development to give greater attention to that connection and to "focus" a "significant" portion of United States foreign assistance on environmental restoration.[5] The impact of that resolution and other such initiatives, whether unilateral, bilateral, or multilateral, will depend on the substance accorded subjective terms like "focus" and "significant."

A nation-state, whether the United States or some other country, provides definitional substance to such terms and approves regulatory regimes in accordance with priorities grounded in some value system. Whether these various initiatives prove helpful ultimately depends to a large degree on the character of that value system which informs a state's implementation efforts.

To the extent that the resolution of environmental problems is successful, the prospects for peace in the world will be enhanced. The key is agreement, agreement on what the problems are and agreement on how they should be resolved. Broad, consistent agreement on environmental matters having a global impact is possible only if basic values are shared, that is only if there is an international environmental ethos. The potential for international conflict in the context of environmental problems increases to the extent that value systems clash or are misunderstood. Conversely, states can minimize the likelihood of conflict to the extent that they arrive at a consensus on the relevant values or further clarify their unique values.[6]

---

3. H.R. Con. Res. 248, 101st Cong., 2d Sess. (1990). *See infra* text accompanying notes 109-25.

4. H.R. Con. Res. 248, 101st Cong., 2d Sess. (1990).

5. *Id. See infra* text accompanying notes 116-17.

6. *See* Tina S. Boradiansky, Comment, *Conflicting Values: The Religious Killing of Federally Protected Wildlife*, 30 NAT. RESOURCES J. 709, 710 & nn.8-9 (1990). *Cf.* Willam Kates, *Indians, Activists Oppose Hydro Project*, BANGOR DAILY NEWS, July 16, 1991 (Native Americans in conflict with Canadian and American utilities over James Bay II reservoir project).

The purpose of this Article is to review three different approaches to resolving global environmental problems peacefully and to make a preliminary attempt at articulating a model international environmental ethos that can be used as a standard for the evaluation of these and other approaches.[7] Having such a model will provide a more objective standard against which proposals and implementation can be measured. Urgency in the need for a response is presupposed.[8]

I. THREE APPROACHES TO INTERNATIONAL ENVIRONMENTAL PROBLEMS

The international dimension of environmental degradation is increasingly prominent. Obviously, pollution and its effects are not respecters of political boundaries. Even localized environmental problems share common characteristics from nation to nation, and to the extent that these problems result in tensions domestically, international stability is threatened.[9] Moreover, there is a growing tendency

---

7. Some would say that the Stockholm Declaration was the first attempt to articulate an international environmental ethos. That effort, however, was more specific than what I am proposing here —going beyond the description of an ethos to the application of an ethos, largely unarticulated, that was more or less assumed. Subsequent efforts, such as the Brundtland Commission Report, have been of a similar character. Although there are similarities in what is being attempted in this Article, I am seeking to back up, to look behind the Declaration and other such agreements, and to ask what should form the foundation for such efforts. Professor Joseph Sax has asked the question in the particular context of rights and "baseline democratic values." Joseph L. Sax, *The Search for Environmental Rights*, 6 J. LAND USE & ENVTL. L. 93 (1990).

8. That things may be more urgent than we think is well put by one commentator:

> Our planet is like a pond, our environmental crises like the lily plant that doubles in size daily. If unchecked, the lily plant will cover the pond in 30 days and choke off all other forms of life. We seem to feel that we need not act until our pond is half covered. When will that be? On the 29th day, of course. We will soon have, figuratively speaking, only one day to save our pond.

Mark Allan Gray, *The United Nations Environment Programme: An Assessment*, 20 ENVTL. L. 291, 318-19 (1990) (paraphrased from Myres S. McDougal & Jan Schneider, *The Protection of the Environment and World Public Order: Some Recent Developments*, 45 MISS. L.J. 1085, 1123-24 (1974)). *See also* Marlise Simons, *Dead Mediterranean Dolphins Give Nations Pause*, N.Y. TIMES (Int'l), Feb. 2, 1992, at 12.

9. *See generally* Philip W. Quigg, *Environment: The Global Issues* (Headline Series, No. 217) (Foreign Pol'y Ass'n Oct. 1973). Commentators acknowledge that domestic conditions may constitute a "threat to the peace" sufficient enough to justify sanctions under the United Nations Charter. *See generally* RICHARD B. LILLICH, INTERNATIONAL HUMAN RIGHTS 442-563 (focusing on economic sanctions against Rhodesia).

to look upon local resources traditionally viewed as subject to the sovereignty of one nation as earth resources belonging to all peoples and therefore properly subject to international regulation.[10]

In response to these issues, and in recognition, at least implicitly, that environmental matters often are matters that affect international relations, a wide spectrum of devices has been, and is being, employed in an effort to address the issues. This Article focuses on three approaches along that spectrum reflected in (1) the Stockholm Declaration, (2) some fifteen wildlife trade agreements identified by the United States International Trade Commission as employing trade restrictions as an implementation tool,[11] and (3) House Concurrent Resolution 248. All three approaches are then evaluated in light of the values they represent and the global attempt to reconcile the frequently opposed interests of preservation and development.

A. Stockholm Declaration

The first approach is the Stockholm Declaration. Representatives from 113 nations attended the United Nations Conference on the Human Environment held in Stockholm, Sweden, in June 1972.[12] As is evident from the Declaration's title and the preamble, people are its primary concern.

Largely hortatory rather than mandatory in its language, and certainly so in practice given its status as a United Nations General Assembly declaration, the Stockholm Declaration lays down twenty-six "principles" to guide the conduct of nations with respect to the "human environment."[13] The Declaration is relatively nonspecific, has no express enforcement mechanism, and reflects the tensions between industrialized nations and developing countries in the context of the Cold War in 1972, the year of the Declaration's adoption.

In Principle 1, the Declaration asserts that there exists a connection between the environment and many of the commonly accepted

---

10. Wildlife is an example of a resource increasingly viewed as having an international character, and the Convention on International Trade in Endangered Species of Wild Fauna and Flora, Mar. 3, 1973, 27 U.S.T. 1089, 993 U.N.T.S. 243 [hereinafter CITES], *see infra* text accompanying notes 75-89, reflects the widespread agreement on the point, at least in the endangered species context. "Wildlife was once considered largely a national or even local asset but is now recognized by an increasing number of countries as an international resource to be conserved for the benefit of everyone." U.S. INT'L TRADE COMM'N, PUB. NO. 2351, INTERNATIONAL AGREEMENTS TO PROTECT THE ENVIRONMENT AND WILDLIFE, 5-29 (1991) [hereinafter USITC 2351].

11. For a description of the 15 agreements identified by the United States International Trade Commission, see USITC 2351, *supra* note 10 and *infra* text accompanying notes 69-108 at 5-1 to 5-2.

12. 1972 U.N.Y.B. 317-37, U.N. Sales No. E.74.I.1.

13. Stockholm Declaration, *supra* note 1, at 1416.

civil and political "human rights."[14] Included with those rights is "the fundamental right to . . . adequate conditions of life, in an environment of a quality that permits a life of dignity and well-being."[15] Also emphasized is the responsibility to consider the welfare of future generations.[16]

Principle 2 affirms the need to protect the "natural resources of the earth" and "especially representative samples of natural ecosystems."[17] The proposed mechanism is "careful planning or management, as appropriate."[18] Principle 3 makes it a priority to maintain and "wherever practicable" to restore or improve the earth's capacity "to produce vital renewable resources."[19]

Under Principle 4, "wildlife and its habitat" and "[n]ature conservation" generally are to be given special consideration in the course of economic development.[20] Principle 5 focuses on the need to protect and share "non-renewable resources."[21]

Principles 6 and 7 have as their concern specified pollution problems. Principle 6 discusses the "discharge of toxic substances" and the "release of heat,"[22] focusing on the protection of ecosystems, and the principle asserts that "[t]he just struggle of the peoples of all countries against pollution should be supported."[23] Principle 7 looks to pollution of the seas, listing as areas of concern "hazards to human health," harm to "living resources and marine life," damage to "amenities," and interference "with other legitimate uses of the sea."[24]

Principles 8 through 15 build on the earlier and more general principles by acknowledging and, to a limited degree, addressing the conflicting economics of development on the one hand and preservation or protection of the natural environment on the other. Principle 8 describes economic and social development as "essential" to enhancing the quality of life.[25] Principle 9 urges the transfer of financial and technological assistance to developing countries to remedy problems presented by "the conditions of underdevelopment

---

14. *See id.* at 1417-18.
15. *Id.* at 1417.
16. *Id.* at 1418.
17. *Id.* The natural resources listed include "air, water, land, flora, and fauna." *Id.*
18. *Id.*
19. *Id.*
20. *Id.*
21. *Id.* Principle 5 urges that non-renewable resources be used so as to prevent their future exhaustion. *Id.*
22. *Id.*
23. *Id.*
24. *Id.*
25. *Id.*

and natural disasters."[26] Principle 10 contends that payment of fair prices to developing countries for raw materials is necessary for their "environmental management" efforts.[27]

Principle 11 declares that environmental policies "should enhance and not adversely affect the . . . development potential of developing countries" and should not hinder the improvement of living conditions for all.[28] "[A]ppropriate steps" should be taken to reach agreement on how to address the economic consequences of environmental measures.[29] Principle 12 again accords developing countries special treatment and generally urges the use of resources to preserve and improve the environment.[30] Principle 13 encourages states to seek compatibility between development and "the need to protect and improve the human environment for the benefit of their population."[31] Principle 14 characterizes "[r]ational planning" as essential to that compatibility effort.[32] Principle 15 singles out towns and cities for special planning attention and seeks the maximization of social and economic, as well as environmental, benefits "for all."[33]

Principles 16 through 20 address implementation. In Principle 16, states are exhorted to apply "[d]emographic policies" as "appropriate" to deal with population growth, "excessive population concentrations," or "low population density."[34] Once again, the twin concerns of the environment and development are identified.[35] Principle 17 urges the delegation of environmental administrative responsibilities to "[a]ppropriate national institutions."[36] In Principle 18, the Declaration recognizes the potentially key role of "[s]cience and technology" in resolving environmental problems, but does so in the context of their more general "contribution to economic and social development."[37] Principle 19 asserts that education is needed and that the media should assist in the education effort.[38] The goal is the protection and improvement of the environment "in its full human dimension . . . in order to enable man to develop in every respect."[39] Principle 20 identifies the need to promote scientific research, devel-

---

26. *Id. Cf.* H.R. Con. Res. 248, *supra* note 4 (arguably consistent with Principle 9 of Stockholm Declaration); *see infra* text accompanying notes 109-25.
27. Stockholm Declaration, *supra* note 1, at 1419.
28. *Id.*
29. *Id.*
30. *Id.*
31. *Id.*
32. *Id.*
33. *Id.*
34. *Id.*
35. *Id.*
36. *Id.*
37. *Id.* at 1420.
38. *Id.*
39. *Id.*

opment, and the dissemination of relevant information and technologies "without constituting an economic burden on the developing countries."[40]

The remaining principles reiterate in an environmental context a number of the traditional international legal norms that found early expression in the United Nations Charter and subsequent actions of international bodies. Principle 21 affirms the proposition that states are sovereign, that they have "the sovereign right to exploit their own resources pursuant to their own environmental policies."[41] That proposition is qualified by the assertion that states at the same time have "the responsibility to ensure that activities within their jurisdiction or control do not cause damage to the environment of other States or of areas beyond the limits of national jurisdiction."[42] Principle 22 addresses remedies for breach of the responsibility identified in Principle 21. States are to cooperate in the development of international law that provides for "liability and compensation for the victims of pollution and other environmental damage."[43]

Principle 23 asserts the need to recognize localized state values as potentially determinative factors in the resolution of environmentally based conflicts.[44] In addition, like many of the other principles, Principle 23 suggests that a double standard, one for developing countries and another for industrialized nations, will be appropriate in some cases.[45] While recognizing national sovereignty, Principle 24 urges cooperation among nations "on an equal footing," specifically referring to the vehicles of bilateral or multilateral agreements.[46] Principle 25 identifies "international organizations" as important players in the environmental arena.[47] Principle 26 asserts that nuclear weapons and "all other means of mass destruction" must be eliminated.[48]

---

40. *Id.*

41. *Id.*

42. *Id. See Trail Smelter Case* (*U.S. v. Canada*), 35 AM. J. INT'L L. 684 (1941). Obviously significant is the potential impact of Principle 21's assertion that, notwithstanding their sovereignty, states must not engage in activities that damage areas beyond their jurisdiction. Except for the possibility that domestic tensions caused by internal environmental harm might spill over into the international arena, *see supra* note 9, Principle 21, if taken to heart, would eliminate virtually all environmentally induced international strife.

43. Stockholm Declaration, *supra* note 1, at 1420. *See also* RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES §§ 601, 602 (1987).

44. Stockholm Declaration, *supra* note 1, at 1420.

45. *See id.* For elaboration of the notion that different standards legitimately may be applied to different countries, see D.B. Magraw, *Legal Treatment of Developing Countries: Differential, Contextual, and Absolute Norms*, 1 COLO. J. INT'L ENVTL. L. & POL'Y 69 (1990).

46. Stockholm Declaration, *supra* note 1, at 1420-21.

47. *Id.* at 1421.

48. *Id.*

The Stockholm Declaration is not and was not intended to be the last word on international environmental concerns. It does not, for example, purport to provide detailed implementation plans or enforcement mechanisms. Moreover, the philosophic source of the enumerated principles is not entirely clear. The twenty-six principles presuppose certain things about people and their relationship to the world and to one another. At times, the principles seem consistent with reality and, thus, adequate. At other times, however, the principles appear to be grounded on presuppositions that are skewed, inconsistent with reality, or incomplete. Consider the preamble.

The introductory clauses emphasize the delegates' concern for identifying shared values. The delegates "considered the need for a *common* outlook and for *common* principles."[49] The entire planet was their priority. They spoke of "the peoples of the world."[50] The delegates' interest was in not only preserving, but also in enhancing the human environment.[51] What "enhancement" of the human environment means, however, is not self-evident.

The major portion of the preamble consists of seven paragraphs of proclamation. In the first paragraph, the Declaration acknowledges the interrelationship between humanity and the natural environment. It acknowledges the power of humanity to alter the natural environment and asserts that both the "natural and man-made" components of the environment are "essential to [a person's] well-being and to the enjoyment of basic human rights."[52] While recognizing that the human environment presents "the opportunity for intellectual, moral, social and spiritual growth,"[53] however, neither the preamble nor the principles that follow explain the extent to which these natural and man-made components are essential or the basis for that essentiality.

In the second proclamation paragraph of the preamble, the Declaration reiterates its focus on the "well-being of peoples," asserting that the "protection and improvement" of the environment affects "economic development" as well.[54] Of course, few could argue that "protection and improvement of the human environment" are not worthy goals. Exactly what that means, however, is another matter. Not surprisingly, the Declaration makes no direct attempt at definition. Declaring that those goals are the "desire" of all peoples and the "duty of all Governments"[55] therefore poses little threat. The second paragraph does use the adjective "urgent."[56] Having

49. *Id.* at 1416 (emphasis added).
50. *Id.*
51. *Id.*
52. *Id.*
53. *Id.*
54. *Id.*
55. *Id.*
56. *Id.*

recognized in the first paragraph the "rapid acceleration of science and technology,"[57] the Declaration early establishes urgency as a characteristic of the needed response to the environmental dilemma.

In the third paragraph, the Declaration describes the capacity of people either to enhance their environment or to harm it. Humanity can use its powers "wisely" or "[w]rongly or heedlessly."[58] Once again, no attempt is made to describe enhancement, but a list of harmful results of ill-conceived applications of humanity's powers is provided: pollution throughout the earth, disruption of ecosystems, depletion of resources, and "gross deficiencies harmful to the physical, mental and social health of man."[59] Perhaps enhancement is nothing more than the reduction or elimination of these detrimental consequences of people's activities.

Paragraph 4 of the preamble's proclamations raises starkly the conflict between "developing" and "industrialized" countries.[60] According to that paragraph, the environmental problems of developing countries consist primarily of the classic human concerns concomitant to a relatively low standard of living. The Declaration concludes therefore that "developing countries must direct their efforts to development."[61] Industrialized countries, on the other hand, "should make efforts to reduce the gap between themselves and the developing countries."[62] Realistically, to accomplish both without further harming the natural environment would seem to require a reduction in the standard of living of the industrialized nations. That politically sensitive proposition is left unstated.

The fifth paragraph is the epitome of diplomatic ambivalence. Noting that population growth is a major problem, the paragraph states that "adequate policies and measures should be adopted, as appropriate."[63] Imagining what is "adequate" or "appropriate" conjures up all sorts of visions. The paragraph goes on, however, to exalt human achievement, describing people as "the most precious . . . [o]f all things in the world."[64] Humanity's capacity "to improve the environment increases with each passing day."[65]

Paragraph 6 seeks to put into the context of time—past, present, and future—humanity's responsibility for its environment. The Declaration here uses the following words to describe those things dependent on the environment to which we should dedicate ourselves:

---

57. *Id.*
58. *Id.*
59. *Id.*
60. *Id.* at 1416-17.
61. *Id.* at 1417.
62. *Id.*
63. *Id.*
64. *Id.*
65. *Id.*

"our life and well-being," "for ourselves and our posterity a better life in an environment more in keeping with human needs and hopes," "the creation of a good life," "freedom in the world of nature," "a better environment," and improvement of the human environment "for present and future generations."[66] The Declaration asserts that along with defending and improving the environment, we should pursue at the same time "the established and fundamental goals of peace and of world-wide economic and social development."[67]

As does the preamble, all of the twenty-six enumerated principles raise significant questions that, for the most part, are left unanswered. For example, regarding Principle 2's affirmation of the need to protect natural resources and ecosystems, the questions "why," "to what degree," and "at what cost" are not addressed. Similarly, how we are to facilitate such protection is an open issue. Why Principle 3 urges maintenance, restoration, and improvement of the earth's capacity to produce renewable resources, as opposed to whatever the alternatives might be, and how we are to decide whether restoration or maintenance is "practicable" are open questions. Considering Principle 6, defining what constitutes a "just struggle . . . against pollution" is a task not attempted by that principle. Likewise, the Declaration has not established what values will identify the "amenities" and truly "legitimate" uses of the sea worthy of protection. And so on.[68]

B. Wildlife Trade Agreements

Treaties and conventions are a second approach to addressing international environmental concerns. Relying on the fundamental

---

66. *Id.*

67. *Id.* For discussion of paragraph 7 of the preamble, see *infra* text accompanying note 130.

68. Professor Louis B. Sohn observed in 1973 that the 1968 Teheran Conference had unanimously proclaimed that the 1948 Universal Declaration of Human Rights "states a common understanding of the peoples of the world . . . and constitutes an obligation for the members of the international community." Louis B. Sohn, *The Stockholm Declaration on the Human Environment*, 14 HARV. INT'L L.J. 423, 515 (1973) (quoting Final Act of the International Conference on Human Rights, Teheran, 1968, U.N. Doc. A/CONF. 32/41 at 4 (1968)). He suggested that in similar fashion the 1972 Stockholm Declaration would "quite likely" in "the not too distant future" represent the world community's perspective on states' legal responsibilities in the environmental arena. *Id.* Professor Sohn may have been slightly over-optimistic. But he was by no means way off the mark. Many of the principles articulated in the Stockholm Declaration have over the last twenty years become to a substantial and increasing degree descriptive of the shared international perspective on environmental priorities. *See* Louis B. Sohn, *"Generally Accepted" International Rules*, 61 WASH. L. REV. 1073, 1078-79 (1986).

international law doctrine of *pacta sunt servanda*,[69] nations have agreed in a variety of specific environmental contexts to limit what would otherwise be prerogatives of their sovereignty. In a number of these situations, trade restrictions are set up as the means of implementation. While generally effective in theory, given the relative willingness of governments to maintain equality in matters of trade, the scope of subject matters with which such agreements are concerned is limited, and the traditionally vexing problem of enforcing international law has diluted the potential effectiveness of such arrangements. Moreover, to the extent that these agreements are ignored by the parties and the agreements' objectives frustrated, the potential for international conflict is arguably greater than would be the case had no agreement been signed.[70]

In response to a request from the United States Senate Finance Committee, the United States International Trade Commission compiled a report entitled "International Agreements to Protect the Environment and Wildlife."[71] The Commission identified nineteen agreements, to the majority of which the United States is a party, that employ trade as an enforcement or implementation mechanism in the effort "to protect natural resources, wildlife, and cultural/historical property."[72] Fifteen of the nineteen agreements concern wildlife (other than fish and whales),[73] and four of the nineteen concern "archaeological, cultural, historical, or natural heritage."[74]

---

69. *Pacta sunt servanda* is the principle of international law that describes the legal obligation of a party to an agreement with another state to perform according to the terms of the agreement. *See* WOLFGANG G. FRIEDMANN ET AL., CASES AND MATERIALS ON INTERNATIONAL LAW 328-29 (1969).

70. With respect to the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), see *supra* note 10 and *infra* text accompanying notes 75-89. Two commentators put it this way:

> The pattern of exceptions and reservations established under CITES has caused the Parties considerable conflict, as the Convention does not provide precise definitions of any of the exemptions, or of how the implementation of these exemptions can be reconciled with the achievement of the primary goals of the treaty. Pressures to allow the exploitation through international trade of national wildlife resources have forced heated debate on the interpretation and implementation of the Convention's sometimes vague provisions.

Laura H. Kosloff & Mark C. Trexler, *The Convention on International Trade in Endangered Species: No Carrot, But Where's the Stick?* 17 ENVTL. L. REP. 10222, 10225 (July 1987) (footnote omitted).

71. *See generally* USITC 2351, *supra* note 10.

72. *Id.* at 5-1.

73. *See infra* text accompanying notes 75-107.

74. Three of the four nonwildlife agreements are bilateral agreements between the United States and the countries of Mexico, Peru, and Guatemala, respectively, concerning the recovery and return of stolen archaeological, historical, and cultural properties. *See infra* note 108. The fourth nonwildlife agreement is the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export, and Transfer of Ownership of Cultural Property. *Id.*

The Convention on International Trade in Endangered Species of Wild Fauna and Flora, usually referred to as CITES, is undoubtedly the best known. It was ratified by the United States on September 13, 1973, and entered into force on July 1, 1975.[75] The Convention seeks to protect species of plants and animals that are threatened with extinction or that may become endangered if their trade continues.[76] The basic mechanism adopted to provide this protection is a system of import and export permits. Scientific and management authorities in each state must approve trade in the species concerned before permits may be issued.

Article VIII of the Convention lays out the primary means of enforcement. According to subparagraph 1,

> [t]he Parties shall take appropriate measures to enforce the provisions of the present Convention and to prohibit trade in specimens in violation thereof. These shall include measures:
> (a) to penalize trade in, or possession of, such specimens, or both; and
> (b) to provide for the confiscation or return to the State of export of such specimens.[77]

Article VIII also establishes parameters for any permitted trade, for confiscation, for recordkeeping, and for reporting.[78]

In the United States, CITES was implemented by the Endangered Species Act of 1973.[79] Primary responsibility for overseeing compliance with CITES in the United States is vested in the Department of Interior's Fish and Wildlife Service.[80] In a letter to the United States International Trade Commission, dated July 25, 1990, from the Office of Management Authority of the Fish and Wildlife Service, the Service outlined several significant problems with the implementation of CITES.[81] According to the Service, the full potential of

---

75. CITES, *supra* note 10.

76. CITES presently protects more than 20,000 plant species and 500 animal species. More than 100 countries are party to the Convention. USITC 2351, *supra* note 10, at 5-29.

77. CITES, *supra* note 10, art. VIII(1), 27 U.S.T. at 1101. Enforcement measures vary widely from nation to nation because of differing perspectives on what is "appropriate." Kosloff & Trexler, *supra* note 70, at 10225 n.38.

78. CITES, *supra* note 10, art. VIII, 27 U.S.T. at 1102-03.

79. 16 U.S.C. §§ 1531-1543 (1988). *See also* Lacey Act Amendments of 1981, 16 U.S.C. §§ 3371-3378 (1988) (providing civil and criminal penalties for foreign law violations relating to wildlife).

80. The Fish and Wildlife Service has promulgated regulations enforcing CITES at 50 C.F.R. § 23.1 (1991).

81. USITC 2351, *supra* note 10, at 5-30, 5-32. For a helpful discussion of the problems presented in the fulfillment of obligations under CITES, see Kosloff & Trexler, *supra* note 70.

CITES in the international arena has not been realized because of limited implementation by many nations.[82]

The Service also noted in its letter that, contrary to the terms of the Convention, trade is occurring with respect to "at least 80 species" for which data as to the effect of their export are insufficient.[83] In addition, the volume of containerized cargo entering the United States makes comprehensive inspection extremely difficult.[84] Funding and staffing in the species' range states often are inadequate for enforcement.[85]

Uncertainty about the efficacy of mutually agreed-to trade restrictions is reflected in the report of the United States International Trade Commission. The Commission made special mention of the African ivory ban under CITES.[86] It noted that five African countries announced their intention to exempt themselves from the constraints of the Convention when the ban became effective in January 1990.[87] The Commission also observed that it had taken testimony critical of trade bans as a means of wildlife protection.[88] Futhermore, the Commission recorded views of other analysts that any success in the preservation of the African elephant would be due not only to a ban on ivory, but also to "a negative stigma" that had attached to the use of ivory as a result of a publicity campaign.[89]

The Convention on Nature Protection and Wildlife Preservation in the Western Hemisphere, also identified in the Commission's report, exemplifies regional agreements to protect wildlife.[90] Entering into force in 1942, the Western Hemisphere Convention was one of several agreements predating CITES that employed trade controls as

---

82. The Fish and Wildlife Service identified trade in sea turtles and elephant ivory as examples of allegedly deficient implementation of CITES.

> The Secretary of the Interior has been asked by several non-governmental organizations to certify under the Pelly Amendment to the Fisherman's Protective Act that nationals of Mexico and Japan (for trade in sea turtles) and of Zimbabwe, China and the United Kingdom (Hong Kong) (for trade in African elephant ivory) have taken actions that diminish the effectiveness of the Convention. The Fish and Wildlife Service is investigating the allegations contained in these requests.

USITC 2351, *supra* note 10, at 5-32.

83. *Id.*

84. *Id.*

85. *Id.*

86. *Id. See generally* M.J. Glennon, *Has International Law Failed the Elephant?* 84 AM J. INT'L L. 1 (1990) (analyzing the impact of CITES on the African elephant).

87. USITC 2351, *supra* note 10, at 5-32.

88. *Id.*

89. *Id.*

90. Convention on Nature Protection and Wildlife Preservation in the Western Hemisphere, Oct. 12, 1940, 56 Stat. 1354, 161 U.N.T.S. 193 [hereinafter Western Hemisphere Convention].

a mechanism of wildlife protection.[91] Like CITES, the Western Hemisphere Convention established a system of export permits as the mechanism by which trade was to be regulated.[92] And as with CITES, implementation has been problematic. In its discussion of the current issues relating to the Western Hemisphere Convention, the United States International Trade Commission identified the northern spotted owl of the Pacific Northwest and marine turtles as species caught in the conflict between economic interests and wildlife protection.[93]

The Commission's report references four bilateral agreements to which the United States is a party that deal specifically with migratory birds.[94] The four conventions variously regulate the taking and trading of specimens of the birds covered, their parts, their eggs, and their nests.[95] In the United States, several statutes have been enacted over the years in response to the nation's obligations under the conventions.[96]

The Agreement on the Conservation of Polar Bears between Canada, Denmark, Norway, the Soviet Union, and the United States,

---

91. In addition to defining trade parameters, the Western Hemisphere Convention, *id.*, obligates parties to set aside territory as wildlife reserves, to restrict hunting, and to enact legislation to protect flora and fauna. USITC 2351, *supra* note 10, at 5-34.

92. Western Hemisphere Convention, *supra* note 90, art. IX, 56 Stat. at 1366-68.

93. USITC 2351, *supra* note 10, at 5-35. *See also infra* note 110.

94. Convention for the Protection of Migratory Birds in the United States and Canada, Aug. 16, 1916, U.S.-Gr. Brit., 39 Stat. 1702 (entered into force Dec. 7, 1916)[hereinafter Canada Convention]; Convention for the Protection of Migratory Birds and Game Mammals, Feb. 7, 1936, U.S.-Mex., 50 Stat. 1311 (entered into force Mar. 15, 1937)[hereinafter Mexico Convention]; Convention for the Protection of Migratory Birds and Birds in Danger of Extinction, and Their Environment, Mar. 4, 1972, U.S.-Japan, 25 U.S.T. 3329 (entered into force Dec. 19, 1974)[hereinafter Japan Convention]; Convention Concerning the Conservation of Migratory Birds and Their Environment, Nov. 19, 1976, U.S.-U.S.S.R., 29 U.S.T. 4647 (entered into force Oct. 13, 1978)[hereinafter Soviet Union Convention]. As can be seen from the title of the Convention with Mexico, it deals with game mammals as well as migratory birds.

95. Trade restrictions are addressed in the Canada Convention, *supra* note 94, art. II, 39 Stat. at 1703; the Mexico Convention, *supra* note 94, arts. II, III, 50 Stat. at 1312-13; the Japan Convention, *supra* note 94, art. III, 25 U.S.T. at 3333; and the Soviet Union Convention, *supra* note 94, art. II, 29 U.S.T. at 4651. Provisions relating to the establishment of refuges and the protection of habitats are also included in the migratory bird conventions. *See, e.g.*, Canada Convention, *supra* note 94, art. IV at 1704; Mexico Convention, *supra* note 94, art. II(b) at 1312; Japan Convention, *supra* note 94, art. VI at 3335; Soviet Union Convention, *supra* note 94, arts. IV, V, VII at 4653-56.

96. *E.g.*, Migratory Bird Treaty Act of July 3, 1918, ch. 128, 40 Stat. 755 (codified as amended at 16 U.S.C. §§ 703-708, 709a-711 (1988)); Migratory Bird Conservation Act of February 18, 1929, ch. 257, 45 Stat. 1222 (codified as amended at 16 U.S.C. §§ 715-715s (1988)).

which entered into force with respect to the United States in 1976, also employs trade restrictions as a conservation device.[97] The taking of polar bears is substantially restricted, and in article V the Agreement requires each party to prohibit export, import, and traffic of polar bears and their parts.[98]

Another agreement noted in the report, the Arrangement Between the United States of America and Canada on Raccoon Dog Importation, entered into force in 1981.[99] The Canadian Ministry of the Environment and the United States Department of the Interior effected the arrangement by an exchange of letters. Under its terms, each party agreed simply to use its "best efforts" to prohibit the importation of raccoon dogs, "a species of wildlife not indigenous to North America [that] threatens an important element of wildlife species indigenous to Canada and the United States."[100]

In its report, the Commission specifically identified five international agreements providing for plant protection through trade regulation. The focus of the agreements is not on the preservation of species, however. Rather, the agreements seek primarily to contain the spread of pests and disease.[101] The Commission also discussed in some detail the International Tropical Timber Agreement, which, while not regulating trade in timber products and therefore not in the Commission's list of nineteen agreements, does provide for monitoring of trade.[102]

Two other wildlife agreements to which the Commission referred and which use trade as a regulatory and enforcement mechanism are

---

97. Agreement on the Conservation of Polar Bears, Nov. 15, 1973, 27 U.S.T. 3918 [hereinafter Polar Bear Agreement].

98. *Id.*

99. Arrangement on Raccoon Dog Importation, Sept. 4, 1981, U.S.-Can., 33 U.S.T. 3764.

100. *Id.* at 3765. The United States International Trade Commission characterizes the arrangement concerning raccoon dogs as an executive agreement that "does not require any implementing legislation." USITC 2351, *supra* note 10, at 5-43.

101. USITC 2351, *supra* note 10, at 5-29, 5-46. The United States International Trade Commission refers to the International Plant Protection Agreement, the North American Plant Protection Agreement, the Plant Protection Agreement for the South East Asia and Pacific Region, the Convention for the Establishment of the European and Mediterranean Plant Protection Organization, and the Phyto-Sanitary Convention for Africa. *Id.* at 5-46.

102. *Id.* at 5-29. More than 40 countries, both consumers and producers of tropical timber, are party to the International Tropical Timber Agreement. *Id.* at 5-45. The United States became a full member of the International Tropical Timber Organization (ITTO) under the agreement in 1990 upon Congress' approval in Public Law 101-246. *Id.* One of the main priorities of the agreement is the development of sustainable tropical timber production. *Id.* at 5-44 to 5-45. Although the agreement was negotiated originally as a "commodity agreement," more recently, in light of deforestation concerns, the "environmental aspect" of the agreement has come to the fore. *Id.* at 5-44.

the African Convention on the Conservation of Nature and Natural Resources and the International Convention for the Protection of Birds.[103] The United States is not a party to either convention, but United States interests are implicated by both.[104] Under the African Convention, the management of wildlife is to occur "with due regard to the best interests of the people" and "within the framework of land-use planning and of economic and social development."[105] The acknowledgment of the development pressures on the nations of Africa is obvious. The International Convention for the Protection of Birds, which entered into force in 1963 and to which ten European countries are party, imposes on the parties a variety of obligations, including restrictions on both international and domestic trade.[106] The Convention, however, contains neither enforcement mechanisms nor reporting requirements.[107]

Finally, the Commission identified four other agreements to which the United States is a party that employ trade restrictions as a means of implementation.[108] Since these agreements deal with archaeological, historical, and cultural property, however, rather than natural environmental matters, the agreements lie beyond the scope of this Article.

C. House Concurrent Resolution 248 and House Bill 3756

A third approach to resolving international environmental problems is unilateral action on the part of a nation. House Concurrent Resolution 248 is one example.[109]

---

103. *Id.* at 5-2.

104. *See id.* at 5-1.

105. African Convention on the Conservation of Nature and Natural Resources, Sept. 15, 1968, 1001 U.N.T.S. 4, 5, 7. Article II of the African Convention is titled "Fundamental Principle" and provides as follows: "The Contracting States shall undertake to adopt the measures necessary to ensure conservation, utilization and development of soil, water, flora and faunal resources in accordance with scientific principles and with due regard to the best interests of the people." *Id.* at 5.

106. International Convention for the Protection of Birds, Oct. 18, 1950, 638 U.N.T.S. 187. *See also* USITC 2351, *supra* note 10, at 5-39, 5-40.

107. USITC 2351, *supra* note 10, at 5-40.

108. *Id.* at 5-2. The four agreements to which the Commission refers are the Treaty of Cooperation Providing for the Recovery and Return of Stolen Archaeological, Historical and Cultural Properties, July 17, 1970, U.S.-Mex. (entered into force 1971); the Agreement for the Recovery and Return of Stolen Archaeological, Historical and Cultural Properties, Sept. 15, 1981, U.S.-Peru (entered into force 1981); the Agreement for the Recovery and Return of Stolen Archaeological, Historical and Cultural Properties, May 21, 1984, U.S.-Guat. (entered into force 1984); and the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export, and Transfer of Ownership of Cultural Property, Nov. 14, 1970 (entered into force 1983). *Id.* at 5-89.

109. H.R. Con. Res. 248, 101st Cong., 2d Sess. (1990).

The status of this resolution in the United States domestic arena is not unlike that of the Stockholm Declaration in the international arena. The House resolution simply articulates the "sense" of the Congress. It is not a statute like the Endangered Species Act and, accordingly, does not require action or forebearance on the part of the executive branch or the people.[110] The resolution is important, though, for at least three reasons. First, it is a valuable reminder of the potential in unilateral action by nations sensitive to international environmental problems. Second, it affirms the potential for influence in foreign aid decisions. Third, it reflects an increasing willingness to acknowledge that environmental concerns present other concerns, specifically concerns about national security, which, given the place of the United States in the world, obviously implicate issues of international conflict and peace.[111] The resolution amounts to an admission that the environmentalists have been right all along on at least one score: environmental irresponsibility has significant externalized costs.[112]

The whereas clauses of the resolution list many of the earth's current environmental problems and connect those problems with

---

110. In contrast to House Concurrent Resolution 248, there have been numerous congressional and administrative agency initiatives in the international natural resources area that have had the force of law. For example, in an effort to protect marine turtles from death due to entanglement in shrimp nets, the United States National Oceanic and Atmospheric Administration's National Marine Fisheries Service requires shrimp harvesters to have on their nets "turtle excluder devices." In its report to the Senate Finance Committee, the United States International Trade Commission noted in addition that importation of shrimp from countries without similar turtle protection measures would be prohibited as of May 1, 1991, under Public Law 101-162. USITC 2351, *supra* note 10, at 5-35. This is an example of one of the most effective methods of resolving international environmental problems and, significantly, it is a unilateral measure. This kind of initiative puts real pressure on foreign traders to conform to higher environmental standards, protects domestic interests from the competition presented by those whose production costs are necessarily lower, and authorizes the government to employ its enforcement powers. In fact, as to shrimp harvesters and turtle conservation efforts, only Surinam shrimp imports were restricted, and that restriction was lifted effective October 1, 1991. Telephone Interview with Alan Risenhoover, Office of Congressional Affairs in National Marine Fisheries Service, Department of Commerce (Feb. 12, 1992). Mexico's imports were not restricted, but instead were made the subject of annual review for three years, the first review being scheduled for May 1, 1992. *Id.*

111. Jessica Tuchman Mathews of the World Resources Institute and former member of the National Security Council has written extensively about the connection between national security and the environment. *See, e.g.*, Jessica Tuchman Mathews, *Chantilly Crossroads*, WASH. POST, Feb. 10, 1991, at C7; Jessica Tuchman Mathews, *Redefining Security*, FOREIGN AFFAIRS, Spring 1989, at 162.

112. *See* Jim MacNeil, *Strategies for Sustainable Economic Development: Balancing Economic Growth and Ecological Capital*, SCI. AM., Sept. 1989, at 154. *See also* Glennon, *supra* note 86, at 6 (African elephant ivory market is a "classic case of market failure").

significant human effects. Characterized as "environmental stresses on the natural resource base," these environmental problems are linked with food shortages and "severe hardships," which in turn will threaten "initiatives to establish healthy, sustainable economies," and "will contribute to increasing political instability, and will constitute a major threat to national security and global peace."[113] "[U]nbridled consumption" in industrialized nations and pollution, deforestation, and "unlimited exploitation of nonrenewable resources" are listed as causes of environmental degradation, which lead to population dislocation and "conflict and tension" as well as "conditions of instability."[114] The preamble to the resolution concludes: "Whereas the geopolitical landscape can change quickly and dramatically due to the political instability resulting from hunger and deprivation brought on by environmental problems: Now, therefore, be it [resolved . . .]."[115]

The resolution urges the Secretary of State and the Administrator of the Agency for International Development to "focus a significant portion of United States foreign assistance on environmental restoration, reforestation, pollution control, family planning[,] improvements in the efficiency of energy use, and rehabilitation of degraded ecosystems."[116] The resolution identifies two primary objectives: the attainment of economies that are sustainable and a reduction in environmentally based political tensions.[117] The means is foreign assistance.

Speaking in support of the resolution at the time of its passage, Congressman Yatron observed that "[n]ational security and global peace are seriously threatened by the continuation of environmental problems."[118] Similarly, according to Congressman Broomfield, "these problems can actually pose threats to international peace and security," and it is "our moral duty" to respond to the environmental problems of the world and the human travails they present.[119] The chief sponsor of the resolution, Congressman Gilman, identified those countries experiencing political unrest due to environmental degra-

---

113. 136 CONG. REC. H7684 (daily ed. Sept. 17, 1990).

114. *Id. See also* Charles E. Di Leva, *Trends in International Environmental Law: A Field With Increasing Influence*, 21 ENVTL. L. REP. 10076, 10081 n.55, 10084-85 (1991).

115. 136 CONG. REC. H7684 (daily ed. Sept. 17, 1990).

116. *Id.*

117. *Id.*

118. *Id.*

119. *Id.* at H7684-85. According to Professor Nancy Lubin of Carnegie-Mellon University, environmental and other severe problems in the central Asian republics of the former Soviet Union, especially Uzbekistan, hold the potential for significant instability within the new Commonwealth of Independent States. National Public Radio broadcast, Dec. 26, 1991.

dation as the ones in special need of foreign assistance.[120] He used the countries of Africa as an example: "The recent destruction of much of Africa's dry land agricultural production was more severe than if an invading army had pursued a scorched-earth policy."[121] Congressman Bereuter emphasized that "unconstrained consumption and waste" in conjunction with the "indiscriminate" exploitation of natural resources "threaten the survival of nations" and operate to the detriment of future generations.[122]

More recently, in November 1991, Congressman Gilman introduced House Bill 3756, which would establish the National Commission on the Environment and National Security to consider the link between the environment and national security. The bill is premised on the proposition that the end of the Cold War and the threat to political stability posed by new and increasing environmental problems call for a reevaluation of national security priorities.[123] Over no more than a two-year period the Commission would "study the changing nature of the national security of the United States in light of recent global political changes and new environmental threats to natural resources, the atmosphere, and ocean resources . . . ."[124] The proposed legislation calls for a preliminary report and a final report from the Commission detailing its findings and making specific recommendations on national security priorities, additional funding, and any necessary institutional changes within the government.[125]

House Bill 3756 is a further acknowledgment that peace and the environment are interrelated and, if enacted into law, holds the potential for valuable first steps in addressing the implications of that connection, at least from the vantage point of the United States. From the perspective of the environmentalists, such a law would be a valuable first step in the internalization of heretofore externalized costs.

---

120. Congressman Gilman introduced an earlier version of Resolution 248 with the following remarks:

> In a number of Third World countries that suffer from deforestation, desertification, soil erosion, and other environmental stresses, People [sic] are faced with chronic food shortages, resulting in political instability. On many occasions, this leads to armed conflicts that involve other nations in terms of emergency relief and military assistance. Such destabilizing actions leave local government prey to hostile takeovers and the geopolitical landscape can drastically shift overnight. In addition to the Third World, the eastern bloc nations struggling with their new found freedoms have suffered from seriously critical environmental problems.

136 CONG. REC. E93 (daily ed. Jan. 30, 1990).

121. 136 CONG. REC. H7685 (daily ed. Sept. 17, 1990).

122. *Id.*

123. 137 CONG. REC. E3833 (daily ed. Nov. 14, 1991).

124. *Id.* at E3834; H.R. 3756, 102d Cong., 1st Sess. § 4(a) (1991).

125. 137 CONG. REC. E3834-35 (daily ed. Nov. 14, 1991); H.R. 3756, 102d Cong., 1st Sess. §§ 4(b), 8 (1991).

D. Summary of Approaches

The Stockholm Declaration is hortatory in character, laying down guiding principles in a multilateral setting. The various international wildlife trade agreements, bilateral and multilateral, affirm the link between the environment and economics and, using trade restrictions as a mechanism of implementation, depend on the parties' adherence to the doctrine of *pacta sunt servanda*[126] for their effectiveness. These agreements reinforce the broader notion that trade, like foreign aid, may be an effective tool in the pursuit of environmental objectives. House Concurrent Resolution 248 and House Bill 3756, both unilateral measures, deal with foreign assistance expressly and foreign policy priorities generally. In doing so, they expressly affirm the premise for this symposium: that there is a link between the environment and political stability and that the prospects for peace in the world are intertwined with the resolution of environmental issues.

II. Values, the Environment, and Peace

A. The Three Approaches

The values a party brings to the interpretation of international environmental agreements determine in great measure the scope and character of that party's compliance. The fact that there often is room for significantly varying interpretation is illustrated by the language of the Polar Bear Agreement. Article II, for example, requires the parties to "take *appropriate* action to *protect* the ecosystems of which polar bears are a part, with *special* attention to habitat components such as denning and feeding sites and migration patterns, and shall *manage* polar bear populations in accordance with *sound* conservation practices based on the *best available* scientific data."[127] The meanings of the italicized words are not self-evident. Although issues of interpretation arise in virtually all legal discourse, the international forum in which such agreements operate, in conjunction with a subject matter about much of which even scientists have not agreed, make for unique difficulties and allow for substantial maneuvering.[128]

---

126. *See supra* note 69.

127. Polar Bear Agreement, *supra* note 97, art. II, at 3921 (emphasis added).

128. Other provisions of the Polar Bear Agreement present the same difficulties. Article VI, for example, expressly requires the enactment and enforcement of such implementing legislation "as may be necessary for the purpose of giving effect to this Agreement." 27 U.S.T. at 3922. The "softness" of the parties' obligations is reflected in the United States International Trade Commission's report to the Senate Finance Committee, in which it noted that increasing oil and minerals exploration in the Arctic and its effect on polar bear habitat are a source of concern. USITC 2351, *supra* note 10, at 5-42.

The central question is: What are the values that all parties share? Consider again the Stockholm Declaration, wildlife trade agreements, and the actions of Congress.[129]

The preamble of the Stockholm Declaration concludes as follows:

> To achieve this environmental goal will demand the acceptance of responsibility by citizens and communities and by enterprises and institutions at every level, all sharing equitably in common efforts. Individuals in all walks of life as well as organizations in many fields, *by their values and the sum of their actions*, will shape the world environment of the future. Local and national governments will bear the greatest burden for large-scale environmental policy and action within their jurisdictions. International co-operation is also needed in order to raise resources to support the developing countries in carrying out their responsibilities in this field. A growing class of environmental problems, because they are regional or global in extent or because they affect the common international realm, will require extensive co-operation among nations and action by international organizations in the common interest. The Conference calls upon Governments and peoples to exert common efforts for the preservation and improvement of the human environment, for the benefit of all the people and for their posterity.[130]

The Declaration here recognizes explicitly the role of values. Governments are identified as primary players, and today, given the global explosion of democracy, these governments will reflect in their policies more than ever before the values and priorities of their peoples. Cooperation regionally and internationally is emphasized. Posterity is a primary focus. In this paragraph, perhaps more than in any other, the Declaration identifies several key values-based issues.

As for the wildlife trade agreements, the preamble of CITES is instructive.

> The Contracting States,
> RECOGNIZING that wild fauna and flora in their many beautiful and varied forms are an irreplaceable part of the natural systems of the earth which must be protected for this and the generations to come;

---

129. These three approaches are connected in an ultimate sense, as is all human endeavor, to a value system of one sort or another. For clarity of analysis, it is helpful to segregate into four components the ways in which such a value system impinge on each approach. First, each approach emanates from, or is motivated by, a certain set of values. Second, values are stated, expressly and impliedly, in the written documents themselves. Third, the means and extent of implementation of each approach by individual parties reflect their individual value systems. Fourth, the assessment of the effectiveness of each approach depends on the value system that informs that judgment.

130. Stockholm Declaration, *supra* note 1, pmbl., para. 7 (emphasis added).

> CONSCIOUS of the ever-growing value of wild fauna and flora from aesthetic, scientific, cultural, recreational and economic points of view;
> RECOGNIZING that peoples and States are and should be the best protectors of their own wild fauna and flora;
> RECOGNIZING, in addition, that international cooperation is essential for the protection of certain species of wild fauna and flora against over-exploitation through international trade;
> CONVINCED of the urgency of taking appropriate measures to this end;
> HAVE AGREED as follows . . . . [131]

The values embodied by CITES and which motivated the drafters and signatories are at least partially apparent from the preamble. Similarly, the preamble of the Western Hemisphere Convention uses words that reveal the values on which that agreement is based and according to which it is to be construed.[132]

The exceptions to the trade restrictions in the migratory bird conventions say something about the trade-offs that the parties are prepared to accept and, consequently, something about their values. For example, the Canada, Japan, and Soviet Union Conventions each permit the taking of birds for subsistence.[133] Those three conventions as well as the Mexico Convention also permit the taking of birds for scientific or educational purposes.[134] Similarly, the Polar Bear Agreement allows for the taking of polar bears for scientific and conservation purposes, by local people and nationals using traditional methods, and "to prevent serious disturbance of the management of other living resources."[135]

Values can change. In commenting on the migratory bird conventions, the United States International Trade Commission in its report to the Senate Finance Committee observed that changes in the

---

131. CITES, *supra* note 10, 27 U.S.T. at 1090.

132. The preamble of the Western Hemisphere Convention reads as follows:

> The governments of the American Republics, wishing to protect and preserve in their natural habitat representatives of all species and genera of their native flora and fauna, including migratory birds, in sufficient numbers and over areas extensive enough to assure them from becoming extinct through any agency within man's control; and
>
> Wishing to protect and preserve scenery of extraordinary beauty, unusual and striking geologic formations, regions and natural objects of aesthetic, historic or scientific value, and areas characterized by primitive conditions in those cases covered by this Convention; and
>
> Wishing to conclude a convention on the protection of nature and the preservation of flora and fauna to effectuate the foregoing purposes, have agreed upon the following Articles. . . .

Western Hemisphere Convention, *supra* note 90.

133. USITC 2351, *supra* note 10, at 5-39.

134. *Id.*

135. Polar Bear Agreement, *supra* note 97, art. III(1), at 3921.

motives for protecting wildlife had occurred over the years:

> The need to protect wild animals has been recognized for many years. However, the reasons for protecting various species have changed significantly. For example, birds were valued originally as a source of food and controller of insects, and early agreements, such as the [Canada Convention], were designed to protect animals useful to agriculture and forestry. Later agreements protected birds because of their aesthetic and cultural value, in addition to other attributes. In recent years, treaties cover a variety of threats and concentrate on habitat protection.[136]

This shift reflects either a change in the values of the parties or the application of old values to new circumstances, or perhaps a combination of the two. Most importantly, however, the world is presently agreed that we ought not to conduct "business as usual." This agreement reflects at a minimum an acknowledgment that the priorities of the past, in certain respects, have been either misaligned or deficient.

Finally, House Concurrent Resolution 248 and House Bill 3756 argue for a redefining of national security. Since World War II most American foreign assistance has been allocated in recognition of the political polarity of the world and, to a large degree, justified in national security terms. These congressional initiatives would seem to assume that the invocation of the term "national security" is essential to the garnering of support for the reallocation of dollars previously earmarked for the Cold War to global environmental problems.

There is undoubtedly a link between political stability and a people's environment, but there are many values other than the maintenance of national integrity that historically have motivated people to act in ways consistent with a national interest. These other values may well suffice to encourage environmental accountability and progress toward sustainable development—as well as foreign assistance or a reallocation of Cold War dollars—without resort to the heretofore talismanic rubric of national security.[137] We should hope so.

B. A Proposed Value System

Having examined three typical approaches to international environmental problems and having identified at least some of the values

---

136. USITC 2351, *supra* note 10, at 5-38.

137. *See* A.J. Fairclough, *Global Environmental and Natural Resource Problems—Their Economic, Political, and Security Implications*, 14 WASH. Q. 78 (1991) (redefining national security and foreign policy in light of environmental problems); Wade Greene, *An Idea Whose Time is Fading*, TIME, May 28, 1990, at 90 (urging a new conception of national security).

that apparently motivated their enactment, one particularly disturbing question remains unanswered. Why does the global environment continue to deteriorate?

Undoubtedly there has been some success. In the United States, for example, the publication of Rachel Carson's *Silent Spring*[138] led to a serious reevaluation of the use of pesticides and herbicides such as DDT.[139] The enactment of the National Environmental Policy Act in 1969[140] marked the beginning of a vigorous two decades of federal regulatory efforts to address environmental degradation. But over all, from a global perspective, during these twenty years since the Stockholm Declaration, we have at best managed to do little more than slow the rate of decline. Obviously, there are many reasons, most of which are interconnected. Let me suggest three.

First, our approach has been piecemeal in character. We have focused on one species or one pollutant. Even schoolchildren know now, however, that ecosystems cannot be compartmentalized.

The second reason is politics. National sovereignty concerns often get in the way of dealing with environmental problems in the uniform and comprehensive fashion necessary for effectiveness.

Third, and probably most significant, is the influence of inappropriate life philosophies. Humanism, pantheism, and hedonism all operate to distort the balance that is essential to abundant life and peace on this planet.[141] Present-day realities reveal both the inadequacy of these philosophies and the need for an articulated consensus on priorities.[142]

In the principles of the Stockholm Declaration, for example, the influence of humanism is apparent and makes the result less than ideal.[143] There is a tendency in the Declaration to suggest that humanity is the measure of all things, that there is no standard higher than humanity. If that is true, then what is determinative is what the majority or the most powerful says. There is no standard to which all people can appeal. There is nothing outside of humanity that gives value to the environment. The environment belongs exclusively to humanity.[144]

---

138. RACHEL CARSON, SILENT SPRING (1962).

139. WALTER A. ROSENBAUM, THE POLITICS OF ENVIRONMENTAL CONCERN 45 (2d ed. 1977).

140. 42 U.S.C. §§ 4321-4370c (1988).

141. Another "ism," materialism, is in many of its applications a subset of both humanism and hedonism, and is therefore not discussed separately here. It is nonetheless similarly opposed to, and destructive of, environmental values generally.

142. *See generally* ZYGMUNT J.B. PLATER ET AL., ENVIRONMENTAL LAW AND POLICY: NATURE, LAW, AND SOCIETY 11-15 (1992).

143. An especially critical parsing of the Stockholm Declaration, decrying its humanistic and pro-development bent, is contained in Mark A. Gray, *The United Nations Environment Programme: An Assessment*, 20 ENVTL. L. 291, 310-12 (1990).

144. The Stockholm Declaration does suggest at least a minimal awareness of

For the most part the Stockholm Declaration avoids the frequent tendency of proponents of environmental protection to equate humanity with the rest of the natural world. Under the pantheists' view, humans are of no greater account than trees. There is a spirit that is common to humans and everything else that we find in the earth—the birds, the flowers, the rocks, and the seas. People deserve no special consideration.

Neither humanism nor pantheism is consistent with the nearly universal view that there is a supreme being who was actively involved in the creation of this planet. Moreover, from the Judeo-Christian standpoint, while it is true that God created man and that man holds a special place in the created order, it is also true that He created everything else that there is.[145] Given humanity's capacity to alter the environment, this places on people a special responsibility. In *Pollution and the Death of Man*, Francis Schaeffer calls it a responsibility of "self-limitation," putting it this way:

> The animal can make no conscious limitation. The cow eats the grass—it has no decision to make; it cannot do otherwise. Its only limitation is the mechanical limitation of its cowness. I who am made in the image of God can make a choice. I am able to do things to nature that I should not do. So I am to put a *self*-limitation on what is possible. The horror and ugliness of modern man in his technology and in his individual life is that he does everything he can do, without limitation. Everything he *can* do he *does*. He kills the world, he kills mankind, and he kills himself.[146]

Other, perhaps less "ultimate," values are implicated by our current dealings with the environment. Stewardship and posterity, for example, are themes common to the Declaration and the other agreements discussed above. There is a declared sense that the current generation will be held accountable for our choices and that future generations are dependent upon us.

As for material objectives, to the degree that development is affirmed as an appropriate end in itself, these and other measures fall short. Development and "progress" have real costs, and there is some development that is simply too costly. The Declaration and other similar affirmations too often assume that planning or man-

---

the eternal, affirming some value in the opportunities for growth in the "moral . . . and spiritual" realms. Stockholm Declaration, *supra* note 1, pmbl., para.1. That the delegates were affirming the existence of a creative, personal God is doubtful, however, given the immediately following reference to "the long and tortuous evolution of the human race on this planet." *Id.*

145. "For by him all things were created . . . ." *Colossians* 1:16 (New Int'l Version).

146. FRANCIS A. SCHAEFFER, POLLUTION AND THE DEATH OF MAN 90 (1970).

agement or technology will provide a safe means to continuing development.[147]

The Declaration suggests other values that should be included in a model environmental ethos. The Declaration identifies aesthetics as a rationale for preservation or protection.[148] It makes human health a focal point,[149] and impliedly acknowledges altruism. The Declaration also exhorts industrialized nations to assist developing countries with their needs. Of course, House Concurrent Resolution 248 is oriented in the same direction, although its concern with foreign aid is tied to United States national security and is not an uninterested preference for the welfare of others.

The Declaration also uses the terms "just" and "fair." While in the Declaration these terms are used only in specific contexts, it would be entirely appropriate for the environmental resolution process to be thought about and pursued within an overarching framework of justice and fairness. Providing definition to those terms admittedly presents its own set of problems, but to acknowledge that those principles are worthy of consideration would be a valuable first step.[150] The same is true of matters of right and wrong. Indeed, that certain things are right and others are wrong is implicit in the preambles of the Declaration and most of the other agreements.

Compared to the Declaration, the agreements identified by the United States International Trade Commission adopt perhaps a more realistic view of human nature and world economic relationships. These agreements take a pragmatic approach, seeking to protect wildlife by removing the financial incentives that unrestricted international trade otherwise affords those for whom the underlying values are not personalized. The approach is piecemeal, however, and enforcement is inconsistent.

Likewise, Congress' recent consideration of the link between national security and international environmental concerns is appro-

---

147. *See generally* John Ntambirweki, *The Developing Countries in the Evolution of an International Environmental Law*, 14 HASTINGS INT'L & COMP. L. REV. 905 (1991) (identifying developing international law themes preferring developing countries). Of course, we must not disavow development altogether while the poverty of the majority of the people in the world cries out for attention. "From a moral perspective, an urgent and undeniable imperative exists that the standard of living of the world's poorest people be improved. Roughly one billion people live in 'absolute poverty,' that is, in a form of existence so characterized by malnutrition, exposure to the elements, disease, and illiteracy that it is below any reasonable standard of human decency." Magraw, *supra* note 45, at 71.

148. In addition to the Stockholm Declaration, a number of the other agreements identify aesthetics as a motivating force. *See, e.g.*, CITES, *supra* note 10, pmbl., 27 U.S.T. at 1090; Western Hemisphere Convention, *supra* note 90, pmbl., 56 Stat. at 1356; Soviet Union Convention, *supra* note 94, pmbl., 29 U.S.T. at 4649.

149. *See generally* Magraw, *supra* note 45.

150. "[T]hese fairness issues will have to be taken seriously." *Id.* at 52-53.

priate to the extent that it acknowledges the ecopolitical realities of our world today and the potential leverage of foreign aid allocations. Of course, at this point, Congress has done no more than declare its collective "sense" of how things are. And, certainly, ensuring national security in a political sense is not the only reason to assist the peoples of other countries.

In any event, humanism and pantheism are the philosophies that more than any others threaten the peaceful resolution of environmental problems. Hedonism also presents a threat and is increasingly accepted as a philosophy of life in more developed countries of the world.[151]

Notwithstanding these threats, as suggested above, there does exist an international environmental ethos of sorts. It is, however, basically unarticulated, incomplete and skewed, and inconsistently applied. Accordingly, what follows is an attempt to articulate a model ethos and to consider the realities of implementation. What is proposed happens to be generally consistent with the value systems of humans through history and across cultures and religions. It is in any event generally inconsistent with humanism, pantheism, and hedonism.

1. Creation.—As suggested above, from the beginning of recorded human history until now, across virtually all civilizations and religions, man has looked to a supreme being as the source and sustainer of life on earth. The Judeo-Christian perspective, based on the biblical account, is one example of that tradition.

"In the beginning God created the heavens and the earth."[152] What follows that first verse in the *Bible* is a detailed account of God's creative work in the world, and the implications are significant. God ordered light, and there was light, and the light was good.[153] He pulled the waters together into seas and ordered the appearance of dry ground, which he called land, and saw that it was good.[154]

> Then God said, "Let the land produce vegetation: seed-bearing plants and trees on the land that bear fruit with seed in it, according to their various kinds." And it was so. The land produced vegetation: plants bearing seed according to their kinds and trees bearing fruit with seed in it according to their kinds. And God saw that it was good.[155]

---

151. Clearly, to the extent that "national security" is a euphemism for preserving the unrestrained pleasure seeking that the American standard of living makes possible, House Concurrent Resolution 248 falls short of being the answer to environmentally based conflict.

152. *Genesis* 1:1 (New Int'l Version).

153. *Id.* 1:3, 4.

154. *Id.* 1:9, 10.

155. *Id.* 1:11, 12.

Following the creation of the stars, sun, and moon,

> God said, "Let the water teem with living creatures, and let birds fly above the earth across the expanse of the sky." So God created the great creatures of the sea and every living and moving thing with which the water teems, according to their kinds, and every winged bird according to its kind. And God saw that it was good. God blessed them and said, "Be fruitful and increase in number and fill the water in the seas, and let the birds increase on the earth."[156]
>
> And God said, "Let the land produce living creatures according to their kinds: livestock, creatures that move along the ground, and wild animals, each according to its kind." And it was so. God made the wild animals according to their kinds, the livestock according to their kinds, and all the creatures that move along the ground according to their kinds. And God saw that it was good.[157]

God, of course, went on to make man, the other side of the environmental equation. He made man in His image, and in so doing established that man was to "rule" over the fish, and the birds, and the livestock, and every creeping thing "and over all the earth."[158] God told man to "[b]e fruitful and increase in number; fill the earth and subdue it. Rule over the fish of the sea and the birds of the air and over every living creature that moves on the ground."[159]

What exactly did God mean in giving man dominion?[160] God explains:

> I give you every seed-bearing plant on the face of the whole earth and every tree that has fruit with seed in it. They will be yours for food. And to all the beasts of the earth and all the birds of the air and all the creatures that move on the ground—everything that has the breath of life in it—I give every green plant for food.[161]

The first chapter of *Genesis* concludes with God's assessment: "God saw all that he had made, and it was very good."[162]

Scripture throughout serves to clarify God's intent for humanity's relationship with His creation. God caused to grow in the garden of Eden "all kinds of trees . . . that were pleasing to the eye and good for food,"[163] and put man in the garden "to work it and take care

---

156. *Id.* 1:20-22.
157. *Id.* 1:24, 25.
158. *Id.* 1:26.
159. *Id.* 1:28.
160. Some have laid blame for the environmental crisis on the wide influence of Christian doctrine. *Cf.* PLATER, *supra* note 142, at 13-14 (considering views of man and nature in various religious traditions). As explained in the text, however, Scripture itself accords the natural world a place of high honor. *See generally* SCHAEFFER, *supra* note 146.
161. *Genesis* 1:29, 30 (New Int'l Version).
162. *Id.* 1:31.
163. *Id.* 2:9.

of it."[164] While He brought the animals to man "to see what he would name them,"[165] they all belonged to Him. Quoting the Lord God, the Psalmist puts it this way: "[E]very animal of the forest is mine, and the cattle on a thousand hills. I know every bird in the mountains, and the creatures of the field are mine. If I were hungry I would not tell you, for the world is mine, and all that is in it."[166] That God has an active interest in the welfare of His created work is evident throughout Scripture.[167] In *Matthew* 10:29, for example, Jesus observes that not a single sparrow falls to the ground without the awareness of God.[168]

Regardless whether the historical Judeo-Christian perspective on the origins of life is accepted, what people across the world generally have in common is the notion that someone or something beyond ourselves has a special interest in and gives value to that which is not man-made.[169] This attribution of value to the natural environment, while not quantifiable, means at the very least that human sensual gratification cannot be the sole measure by which the environmental impact of human activity is evaluated.[170] On the other

---

164. *Id.* 2:15.

165. *Id.* 2:19.

166. *Psalm* 50:10-12 (New Int'l Version). *See also Psalm* 104.

167. The entry of sin into the world through the fall of man had consequences not only for the progeny of Adam and Eve, but also for the rest of creation. *Genesis* 3 records the curse on the ground, the expulsion of man from the garden, and his relegation to a life of working for food. *Genesis* 3:17-19, 22-24. Paul describes the bondage in which we find creation and the restoration for which creation waits:

> . . . The creation waits in eager expectation for the sons of God to be revealed. For the creation was subjected to frustration, not by its own choice, but by the will of the one who subjected it, in hope that the creation itself will be liberated from its bondage to decay and brought into the glorious freedom of the children of God.
>
> We know that the whole creation has been groaning as in the pains of childbirth right up to the present time.

Romans 8:19-22 (New Int'l Version).

168. Importantly, at the same time, contrary to those who would elevate sparrows above humans, Jesus then exhorts His disciples not to fear for their souls since "you are worth more than many sparrows." *Matthew* 10:31 (New Int'l Version). Moreover, *Romans* 1:25 points out the heresy in worshipping and serving the creation rather than the Creator.

169. *See generally* C.S. LEWIS, THE ABOLITION OF MAN (Macmillan Paperback ed. 1965) (1947). *See also infra* text accompanying notes 184-88.

170. *See supra* text accompanying note 146. The Statement of Purpose of Covenant College affirms man's responsibility:

> [W]e seek . . . to accomplish the following general aims in every area of life: (1) to see creation as the handiwork of God and to study it with wonder and respect; (2) to acknowledge the fallen nature of ourselves and of the rest of creation and to respond . . .; (3) to reclaim the creation for

hand, given the sanctity of life as affirmed across cultures through time, the propriety of human activity cannot be assessed from a purely environmental standpoint, irrespective of its impact on human life.

Needless to say, there is a great deal of room between sensual gratification at one extreme and survival at the other, but these two poles set the outer boundaries within which judgments must be made. Other values propositions are necessary to guide the judgments within these parameters.

2. Stewardship.—Where a thing has value and a person is given authority over that thing by the one attributing value to it, the person acts in the role of a steward. A steward is one charged with supervising a matter or directing the affairs of some institution. Stewards are accountable to the ones who place them in control for the results of their stewardship.

As noted above, the Judeo-Christian tradition expressly supports the proposition that God has attributed value to the natural environment. Within this tradition, then, it takes no great leap of faith to conclude that God has given humanity stewardship responsibilities vis-à-vis the natural environment.

Nonetheless, the most strident voices of the environmentalism movement sound in stewardship terms without even pretending to have a faith-based rationale. Stewardship is a concept familiar to all, religious zealot and atheist alike, and in the environmental context has a longstanding history. John Muir and Teddy Roosevelt were stewards of the environment long before the high school students who visited state legislatures on the first Earth Day in 1970. Even if they did not preach the virtue of stewardship, many before Muir and Roosevelt, such as the American Indian, practiced stewardship of the earth's resources. Elsewhere in the world, outside North America, cultures have long sought a harmony with the natural environment that reflects respect and need, and recognizes as well the natural world's intrinsic value.

Stewardship as a precept of life is made manifest to anyone who has owned land.[171] Realizing that a deed is nothing more than a piece of paper, that the land described therein cannot be transported, and that the land will remain long after the owner dies, makes it almost

---

God and to redirect it to the service of God and humankind . . . .

That humanity's dealings with nature raise important moral issues is by no means, however, a proposition that has been voiced solely by those standing on a traditional religious platform. *See, e.g.*, R.L. Means, *Why Worry About Nature?* SATURDAY REV., Dec. 2, 1967, at 13-15 (*reprinted in* SCHAEFFER, *supra* note 146, at 124-25).

171. Consider the historical development of the English common law of property, which focused on the incidents of tenure rather than "ownership." *See generally* S.F.C. MILSOM, HISTORICAL FOUNDATIONS OF THE COMMON LAW (2d ed. 1981).

farcical for one to say that he owns the land.[172] Considering this, ownership really means only that under the law the owner has the exclusive right of possession of that tiny piece of the earth for a short period of time.

3. Health.—Throughout history and across cultures, people have sought good health and long life. Most would agree that these interests spring from an instinct for survival. There are, of course, aberrations, both noble and pernicious. Sacrifice and suicide come to mind. But generally, the clinging tenacity with which individuals seek to survive is a given.

Whether survival constitutes the ultimate, all-encompassing value probably depends on one's theological bent and one's view of "non-life"[173] or "the afterlife," but one would rarely if ever contend that the pleasure or comfort of others should be raised above one's own health or survival. One would be even less likely to argue that technology should be pursued for its own sake regardless of the ramifications for human health and survival. The pleasure-pain utilitarian principle may reflect an individual's philosophy of life, but mankind has not even approached the point of permitting such a philosophy to dictate the priorities of society in general when to do so would compromise the biological well-being of others.[174]

Our common priority of health and survival has already manifested itself in concrete fashion in a variety of contexts. Recognizing the need to preserve genetic diversity in grains to ensure sufficient world food resources, international gene banks have been established. The ozone protocols are motivated by internationally shared concerns for the deleterious health effects of increased ultraviolet radiation.[175]

4. Posterity—Related to the notion of stewardship is the notion that we who are living today bear some responsibility for the state

---

172. *See* SCHAEFFER, *supra* note 146, at 70. "[W]e are to exercise our dominion over these things not as though entitled to exploit them, but as things borrowed or held in trust, which we are to use realizing that they are not ours intrinsically." *Id.*

173. *Cf.* Azzolino v. Dingfelder, 337 S.E.2d 528 (N.C. 1985), *cert. denied*, 479 U.S. 835 (1986) (rejecting "wrongful life" claim); Procanik v. Cillo, 478 A.2d 755 (N.J. 1984) (allowing recovery of extraordinary medical expenses in "wrongful life" case); Becker v. Schwartz, 386 N.E.2d 807 (N.Y. 1978) (law is ill-equipped to compare "life in an impaired state and nonexistence;" it is "a mystery more properly to be left to the philosophers and the theologians.").

174. The relatively recent legalization of abortion and current discussions about euthanasia arguably are exceptions to the proposition that humanity has never seriously considered the adoption of the utilitarian's philosophy.

175. The refusal of countries like China, India, and Brazil to sign the Montreal Protocol on Substances that Deplete the Ozone Layer because they cannot afford to switch to substitutes reflects the tensions that arise when developing countries seek to develop in ways pursued earlier by developed countries. *See Resource Use Tops '92 Agenda*, NAT. L.J., May 13, 1991, at 19.

of the earth occupied by future generations.[176] "We have not inherited the earth from our parents, we have borrowed it from our children," so the saying goes. There is simply no way to get around the fact that what we do to the earth today will affect those who follow us.[177]

For purposes of applying this principle in assessing the acceptability of environmentally limiting activity, we would do well to focus on our personal posterity. The parenting inclination is undoubtedly one of the strongest human inclinations. The well-being of our children is a motivating force on Little League baseball fields in middle America and in the starvation deserts of Ethiopia and Somalia. For many, it is *the* motivating force of their lives. We desire and seek the best for our children. Certainly there is common ground here. To the extent that the environmental impact of human activity affects the present or future well-being of our children, we have a standard on which virtually all can agree.

The trick, of course, is to get each of us to acknowledge that another's children are as worthy of consideration as our own. That will lead to a subordination of our own children's pleasure to the interests of another's children when those interests are of a higher order. For example, when my activity on behalf of my children's comfort or pleasure threatens the survival or health of someone else's children, I should refrain.[178]

5. Material Objectives.—Once we have met the challenges of procuring our own survival, good health, and the well-being of our children, which most of us in the United States have done to a large degree, and perhaps securing the same for our neighbors (near and far), we tend then to think in terms of comfort, enjoyment, satisfaction, and fulfillment. In what direction do these thoughts lead us? Inevitably, we head toward either the natural world or personal relationships, or both. When we travel, we send postcards or buy books depicting the natural beauty of the places we visit. We decorate our homes and offices with live plants and painted landscapes. And even when personal relationships are our immediate direction, we often see their most pleasant long-term expression in a natural pastoral setting.

Similarly, most people have long ago discovered that consumption does not fully satisfy. While even the natural environment does not

---

176. *See* Sax, *supra* note 7 (construing C.S. Lewis' *The Abolition of Man* to support "patrimonial responsibility as a public duty").

177. Indeed, legal rules impose on us the obligation to consider later generations. Consider the Rule Against Perpetuities. LEWIS M. SIMES, HANDBOOK OF THE LAW OF FUTURE INTERESTS 255 (2d ed. 1966).

178. Children are more susceptible to the negative effects of pollution than adults. *Earth Almanac*, NAT'L GEOGRAPHIC, Dec. 1990, at 146 (estimated 14 million children under age five die annually).

meet ultimately and completely the spiritual yearnings of people, few would disagree that on earth the context in which peace of mind and heart is most readily available and sought is a natural context.[179]

All of this goes to the point that it makes no sense to desire and strive for that which is lost in the striving. It is foolish to build for ourselves edifices or institutions or equipment of pleasure that in their construction or operation maim or destroy that which was the reason for their creation. Examples are all around. We like houses in the woods. We go to the woods and build houses. The woods disappear to make places for the houses and to provide construction materials. We like the solitude of the seashore. We go to the seashore and build condominiums. With the erection of the condominiums the solitude disappears. We like ivory jewelry. We go to the African plains and kill the elephants for their tusks. With the killing of the elephants, the ivory disappears. We like fresh air. We buy cars to escape the air in the city which our cars have polluted, and the polluted air follows us in our cars to the mountains. We like the stillness of the open sky. We build jets to take us to meetings to discuss houses and condominiums and skyscraper construction and jewelry imports and car manufacturing, and to get us to places where the sky is open and still. The stillness of the open sky disappears.

Numerous other examples of a more complicated character could be listed, and they are at least as compelling. For example, our thirst in this country for the comforts of mobility and synthetic fabrics spurs the development of oil fields that disrupt the tundra and wildlife of the Arctic, encourages the use of oil tankers in the oceans that lead to oil spills and the destruction of our seashores, and demands the burning of fuel in our cars which in turn heats the atmosphere and alters the climate.

These things, of course, have their value. What the examples point out, however, are the externalities that we so rarely consider in our drive for development. They suggest the importance of considering the broader ramifications of "progress" and whether a particular step of progress means, in fact, a greater distancing from the things we are really pursuing. Developing nations are in a good position to learn from the mistakes of the "developed" nations in this regard. And to the extent that the environmental effects of maintaining what "progress" has provided cross national boundaries, developed nations must be prepared to retrace some of their steps if they expect developing nations to forego the "progress" that now is

179. Thoreau expressed it often and well. "I went to the woods because I wished to live deliberately, to front only the essential facts of life, and see if I could not learn what it had to teach, and not, when I came to die, discover that I had not lived." HENRY DAVID THOREAU, WALDEN 100-01 (1906).

known to threaten the environment. The recent international efforts regarding the chlorofluorocarbon (CFC) threat to the ozone layer are an example of the cooperation and retraction that will be necessary.[180]

6. Beauty.—There are many features of our natural world that are universally acknowledged as beautiful. We share across cultures a sense of awe at the beauty found in the natural environment, from the sparkling of constellations to waterfalls to the flying formations of migrating geese to seas of prairie grass to the symmetry of plankton. What is good should be preserved. Beauty is good. Beauty should be preserved.[181] The syllogism seems instinctive. It provides yet another standard to which we can point in agreement when making internationally significant environmental judgments.

7. Altruism.—It would be inappropriate to assume that everyone acts out of short-term self-interest all of the time. Clearly, that is not the case. Most people are willing on occasion to prefer the interests of others to their own. It happens most frequently in family relationships, but it also occurs in broader contexts. Community-based relationships commonly underlie self-sacrifice, and we call it patriotism when personal interests are set aside in a national context.

The motivations for altruisitic behavior undoubtedly vary. Christians might point to Jesus' account of the Samaritan on the road to Jericho.[182] Others might look to Gandhi or the "New Age" philosophy of Shirley MacLaine, and some would cite Ayn Rand[183] or Machiavelli, discounting all together the characterization of certain behavior as altruistic. On the international front, foreign aid might well find support or explanation in each of these perspectives. Regardless of the motivations, however, there obviously are times that we prefer others' interests to our own.

Whether for reasons of charity or longer-term self-interest, sacrifice in the environmental arena is consistent with the behavioral heritage of humanity in many other contexts. Present-day communications technology makes the world more like the communities of days past, and it is not unreasonable to assume that certain individuals, groups, and perhaps even countries will be willing to respond to more global problems as neighbors have in community settings. Indeed, organizations like World Vision and International Red Cross

---

180. In light of new data indicating that ozone depletion is occurring more rapidly than previously thought, President Bush announced in February 1992 that the timetable for phase-out of CFC production would be advanced to December 31, 1995, four years earlier than treaty obligations presently require. Rose Gutfeld, *U.S. to Step Up Bid to Protect Ozone Layer*, WALL ST. J., Feb. 12, 1992, at A3.

181. SCHAEFFER, *supra* note 146, at 54, 73.

182. *Luke* 10:29-37. The *Bible* says much about altruism. Servanthood, humility, and gentleness are key words in the life of a Christian. *See, e.g.*, *Matthew* 5:5; *Philippians* 2:3-8.

183. *E.g.*, AYN RAND, THE FOUNTAINHEAD (1943).

are vital, well-established models, and their names alone make the point. Similarly, on a more personal level, if we knew, for example, that driving our cars caused starvation of children in East Africa, some of us would consider car pooling or public transportation more seriously. In economic terms, it is to a great degree just a matter of becoming more educated about the externalities.

8. Justice.—Justice is universally accepted as an essential characteristic of right relationships between people. Its meaning has been the subject of philosophical contemplation for ages. Although the many faces of justice render broad definition difficult, most people would agree that there is a right way and a wrong way to treat fellow human beings in particular circumstances. C.S. Lewis maintains that what is right and wrong varies little from culture to culture and across time.[184]

> I know that some people say the idea of a Law of Nature or decent behaviour known to all men is unsound, because different civilizations and different ages have had quite different moralities.
>
> But this is not true. There have been differences between their moralities, but these have never amounted to anything like a total difference. If anyone will take the trouble to compare the moral teaching of, say, the ancient Egyptians, Babylonians, Hindus, Chinese, Greeks and Romans, what will really strike him will be how very like they are to each other and to our own.[185]

Scripture supports Lewis' position. *Romans* 1, for example, teaches that certain norms are embedded in the structure of the universe.

---

184. In *The Abolition of Man*, C.S. Lewis refers to this context for living as the Tao. What is common to the many forms of this concept, whether "Platonic, Aristotelian, Stoic, Christian, [or] Oriental," is "the doctrine of objective value, the belief that certain attitudes are really true, and others really false, to the kind of thing the universe is and the kind of things we are." LEWIS, *supra* note 169, at 28-29. Aberrations have existed, of course. But they are aberrations flowing from aberrational humans who, because we recognize their moral independence, are held accountable for deviating from the standard. Lewis also makes the case against relativism and skepticism as descriptive of the way things are or should be. *See generally id.*

185. C.S. LEWIS, MERE CHRISTIANITY 17 (Fount Paperbacks ed. 1977) (1952). Lewis continues:

> I need only ask the reader to think what a totally different morality would mean. Think of a country where people were admired for running away in battle, or where a man felt proud of double-crossing all the people who had been kindest to him. You might just as well try to imagine a country where two and two made five. Men have differed as regards what people you ought to be unselfish to—whether it was only your own family, or your fellow countrymen, or every one. But they have always agreed that you ought not to put yourself first. Selfishness has never been admired. Men have differed as to whether you should have one wife or four. But they have always agreed that you must not simply have any woman you liked.

Id. at 17-18.

Paul writes that creation itself speaks out what is right and wrong and that men hear by way of their consciences, choosing to listen or to disregard.[186] Similarly, Scripture provides specific normative principles. For example, we ought not to take that which does not belong to us.[187] Truth is right, falsehood is wrong.[188]

In the environmental field, we will make progress to the extent that we insist on adherence to these norms. In fact, the *Restatement (3d) of Foreign Relations*, in Part VI, has provided for specific application of these principles in the international environmental arena. According to section 601, states are obliged to avoid injuries to the environments of other states, and, under section 602, required to pay compensation for any such injuries.[189]

9. Peace.—That peace is a value appropriately shared by the people of the world needs no rationalization or explanation. The United Nations Charter has peace as its overarching objective and fundamental premise.[190] Jesus Himself said, "Blessed are the peacemakers, for they will be called sons of God."[191]

C. Threats to Peace

Adherence to normative principles is a two-way street. Given the lack of world government, one nation cannot expect adherence by another without itself being willing to adhere. And too often, historically, distinctions allegedly justifying different treatment have amounted to nothing more than hollow rationalizations.[192]

Without world government and global enforcement mechanisms, therefore, shared values are essential to an effective international environmental legal regimen. One could also make the case that shared values would be necessary for an effective response to the

---

186. *Romans* 1:18-25.

187. *Exodus* 20:15.

188. *Id.* 20:16.

189. RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW OF THE UNITED STATES §§ 601, 602 (1987).

190. Article 1(1) of the United Nations Charter establishes unequivocally that peace is the priority of priorities:

> The Purposes of the United Nations are:
>
> 1. To maintain international peace and security, and to that end: to take effective collective measures for the prevention and removal of threats to the peace, and for the suppression of acts of aggression or other breaches of the peace, and to bring about by peaceful means, and in conformity with the principles of justice and international law, adjustment or settlement of international disputes or situations which might lead to a breach of the peace.

U.N. CHARTER art. 1(1).

191. *Matthew* 5:9 (New Int'l Version).

192. *Cf.* Jessica Mathews, *Gorilla in the Greenhouse*, WASH. POST, July 25, 1991, at A17 (United States recalcitrance on global warming issues).

environmental crisis even if world government existed. Stability depends on shared values. In either event, disagreement on the above-enumerated nine principles poses two kinds of threats.

The first is the threat that we will disagree on the importance of each of the values listed. Assuming disagreement follows good-faith resort to a relatively objective standard,[193] however, we probably do not need to be too concerned. We probably will be close enough that compromise and consideration of creative alternatives in light of the objectives of each party will be feasible and will present little threat to the environment. The important point here is seeing that influences outside the ethos are not allowed to compromise the acknowledged priorities. "Progress," for example, by itself, is not included in the value system. To whatever extent progress or development is inconsistent with any of the listed values, it must be prohibited.

The second and more serious threat is that a state might refuse as a first step to acknowledge in any honest fashion the primacy of the values listed. This could occur in a couple of ways. One possibility is that one or a few disagreeing individuals with sufficient political or economic power might exercise their power in purely self-interested terms, disregarding the common values, to the detriment of the environment and the health of the rest of us. Saddam Hussein's tactics in the Gulf War will undoubtedly become a classic example.

The more likely possibility, however, is that group interests within a state will predominate because of ignorance or inertia. People who are uninformed have little desire or incentive to act in a way that compromises their short-term comfort. To the extent that Americans or Brazilians, or fur traders or oil company executives, or homemakers or farmers lack information on the impact of their activities, they will have no reason to adopt new priorities or rearrange old ones.[194]

A lack of information, however, is relatively easy to remedy. Inertia or complacency is the bigger concern. Recognizing that each of us is only one of 4.5 billion people on this planet, we are not inclined to think that a change in our everyday activities or even in the way we vote will have any remotely significant effect. The result is that we tend to go on as we have been taught and in the way to which we have become accustomed and that makes us comfortable. While adhering to the above values in our own little world, we will

---

193. *See generally* ROGER FISHER & WILLIAM URY, GETTING TO YES (1981) (guide to successful negotiations).

194. In the corporate boardroom, lack of information is less and less the problem. The "Valdez Principles" promulgated by the Coalition for Environmentally Responsible Economics is representative of the increasing tendency of corporations to adopt environmental ethics standards. *See Business Bulletin*, WALL ST. J., Feb. 13, 1992, at A1.

fail to insist that our society conform. The consequence of such a perspective is a community or national policy that reflects this general apathy, and the self-interest of the loudest groups will predominate in policymaking. To the extent that group interests set the agenda, the broader global environmental interest will likely be seriously compromised.[195]

III. ADDITIONAL APPROACHES

It is probably inevitable that there will be some disagreement on specific application of the values identified above. The threat to peace is directly proportional to the significance of the disagreement. The effect of disagreement can be assuaged, however, by incorporating the values discussed above into general legal obligations that are more binding in their character than international or congressional declarations or even the wildlife trade agreements discussed earlier, and by providing for obligatory dispute resolution mechanisms. What must be minimized is the opportunity for states to avoid the application of previously agreed-to principles when unanticipated matters threatening a state's short-term interests arise.

In the United States, self-executing treaties would go far in avoiding the negative aspects presented by the fickleness of daily politics.[196] The apparent success of the individual petition mechanism under the European Convention for the Protection of Human Rights and Fundamental Freedoms suggests another possibility for the international environmental arena.[197] Using the model offered by citizen suit provisions in federal pollution control statutes,[198] enforcement then would not be solely dependent on states or international organizations. Both of these mechanisms would tend to assure greater access to the American court system than is ordinarily available on international matters.

The Permanent Court of Arbitration at The Hague, identified as the forum for dispute resolution under a number of international

---

195. There is a third possible threat. That is the threat that people of influence, so-called "intellectuals" perhaps, will persuade us that values are relative. LEWIS, *supra* note 169, at 34-35. If that happens, we may find it extremely difficult to agree that the values listed above are shared or should be shared. We then are without a foundation on which to build a responsive global environmentally sensitive matrix. Those who decide that there are no absolutes will conclude that neither the environment nor anything else is worth saving. They inevitably will come into conflict with those who decide otherwise.

196. *See generally* RICHARD B. LILLICH, INTERNATIONAL HUMAN RIGHTS 91-117 (1991).

197. *See id.* at 642-46.

198. *See generally* Michael S. Greve, *The Private Enforcement of Environmental Law*, 65 TUL. L. Rev. 339 (1990).

agreements,[199] or similar international tribunals might be vested with wider-ranging jurisdiction to resolve environmentally based disputes.[200] This would assure a forum of a less adversarial character and over time could lead to the development of "case law" with its concomitant advantages of predictability and clarification of obligations. The option of economic sanctions for environmental irresponsibility should be explored further. To the extent that political conflict between states is the result of environmentally significant activities, the United Nations Charter already provides a mechanism for securing sanctions,[201] and the effectiveness of sanctions is increasingly apparent and acknowledged.[202]

IV. PRACTICAL POINTS OF COOPERATION

A. Scientific Research

"Two heads are better than one." Cooperative projects and an organized sharing of results would enhance efficiency and increase speed in responding to common environmental problems. Joint governmental projects as well as joint projects sponsored by non-governmental organizations are worthy candidates. University-based cooperative efforts are likewise potentially fruitful options that should be encouraged. A global focus or a regional focus, depending on the problem being addressed, are both appropriate.[203]

B. Parallel Legal Regimes

Through the European Convention on Human Rights, the countries of Europe have achieved some success in establishing consistent

---

199. *See, e.g.*, CITES, *supra* note 10, art. XVIII, 27 U.S.T. at 1114.

200. *Cf.* Stephen M. Schwebel, *Reflections on the Role of the International Court of Justice*, 61 WASH. L. REV. 1061 (1986) (highlighting weaknesses of International Court of Justice).

201. U.N. CHARTER arts. 39, 41.

202. *See* Valeria N. Spencer, Comment, *Domestic Enforcement of International Law: The International Convention for the Regulation of Whaling*, 2 COLO. J. INT'L ENVTL. L. & POL'Y 109 (1991) (asserting advantages of economic sanctions and private pressure in light of "weak response" of United States under Pelly and Packwood-Magnuson Amendments). In the face of threatened economic sanctions, Japan decided in November 1991 to ban drift net fishing by the end of 1992, expressly acknowledging the role of international pressure. *See, e.g.*, *Dolphins' Day: Japan*, ECONOMIST, Nov. 30, 1991, at 34.

203. The Japan and Soviet Union Migratory Bird Conventions, *supra* note 94, provide for joint research programs in articles V and IV, respectively. Likewise, the Polar Bear Agreement, *supra* note 97, in article VII calls for coordinated research, consultation, and exchange of information. Polar Bear Agreement, *supra* note 97, 27 U.S.T. at 3922.

legal regimes in the human rights field.[204] In the environmental area, by treaty, countries might likewise choose to approach common environmental problems in parallel fashion.

Global problems require global responses, and generally the more consistent the response the more effective. Without a dependable international enforcement mechanism, however, the international community is dependent on the law enforcement mechanisms of its member states. Mutual treaty standards and obligations reduced to national law gain the enforcement status of "real" law and thereby move environmental responses from aspiration to enforceable obligation.[205]

Perhaps the most fruitful area of endeavor today would be increased joint efforts to internalize the heretofore externalized environmental costs of engaging in certain businesses. The so-called realities of the day demand that market competition be given great weight in the equation that determines the burdens to be imposed on business. President Bush's trip to Japan highlights the significance of this consideration. Unless two countries agree that each will impose equal burdens on its companies in their attempts to internalize environmental costs, the self-interested trade priorities of individual states may well dictate a particular nation's response.[206]

C. Education

The people of the world need to be educated about environmental causes and effects. Many, of course, experience the effects firsthand. Virtually everyone, however, could benefit from a better understanding of the broader ecological effects of certain lifestyles.

The activities of the United States Fish and Wildlife Service under the Western Hemisphere Convention illustrate the kind of cooperative education programs that might be implemented. For example, the United States has jointly sponsored graduate wildlife management programs at universities in Costa Rica and Brazil, as well as an international training center for managers of reserves in Mexico.[207]

---

204. *See* LILLICH, *supra* note 196, at 678-79 (letter to THE TIMES (London), Mar. 4, 1978, regarding "domestication" of European Convention on Human Rights).

205. Mutuality of obligations ordinarily avoids the types of arguments that spring from "equal protection" concerns. For example, "[i]mport and export permit requirements do not directly interfere with a nation's internal affairs, *nominally affect all signatories equally*, and tend to affect relatively small political interest groups, such as wildlife traders." Kosloff & Trexler, *supra* note 70, at 10226 (emphasis added).

206. The nineteen agreements identified by the United States International Trade Commission are a start. *See supra* notes 71-108 and accompanying text; *see also supra* note 110.

207. USITC 2351, *supra* note 10, at 5-35.

D. Free Flow of Information

The recent collapse of many of the political and economic ideological barriers in the world has served to open channels of information that can be used to the great advantage of all in our need to respond to the environmental concerns of our day. These channels must remain open if we are to maximize the potential inherent in cooperative scientific enterprises and heightened environmental awareness on the part of the world's people.[208] In addition, the free flow of information allows conversation on the more basic questions of values that must undergird more specific policy-oriented discussions about environmental concerns. The relatively closed channels in the Middle East and China presently inhibit those important preliminary interactions.

CONCLUSION

I have attempted to begin articulating a standard by which the world community can measure its environmental proposals. This standard is derived from what appear to be shared values. Assuming that these values are in fact shared across cultures and political boundaries, recognizing that truth should help us objectify our discussions and negotiations.

Irrationality and misunderstanding lead to tension and distrust. To the extent that these interactions are objectified, we insert into the environmental crisis resolution process a rationality that reduces tensions and promotes peace. Perhaps more importantly, the values listed respond to the destructive human tendencies toward greed and ambition. Openly identifying and acknowledging these values as primary for all peoples encourages behavior consistent with those values and is a major step toward mutual accountability. International peace is necessarily promoted.

---

208. *See* Sax, *supra* note 7. Indeed, some have maintained that cooperation in the prevention of pollution is required by customary international law. *See, e.g.*, Magraw, *supra* note 45, at 86-87 n.71.

Outer Space and Peace: Some Thoughts on Structures and Relations

GLENN H. REYNOLDS*

As the last few years of the twentieth century run by, it is natural to begin thinking of what challenges and opportunities the next century will bring. In terms of challenge and opportunity, one area of human activity that promises to be very significant is the exploration and development of outer space. Like other frontiers, from the American West to the deep seabeds to the frozen (but not barren) shores of Antarctica, outer space promises both risks and rewards.

And, as with other frontiers, different commentators have viewed the promises and perils of the space frontier differently. Some have feared that outer space might become simply a new arena for old, formerly earthbound conflicts and a destablilizing force with regard to the situation on Earth. Others, such as the turn-of-the-century Russian space writer Konstantin Tsiolkovsky, have hoped that the expansion of humanity into outer space might usher in "the perfection of human society and its individual members" by providing access to the almost limitless material wealth of the Solar System.[1]

Either scenario is possible, but what actually happens will depend a great deal on what we do at the outset. Although recessions and arms races, technical breakthroughs or failures, and political will or timidity may cause the process to speed up or slow down, the ultimate outcome likely will be the expansion of humanity beyond the surface of the Earth and throughout the Solar System. As Daniel Boorstin has observed, technology is a field in which there are neither counterrevolutions nor restorations.[2]

But although the outcome itself may be more-or-less inevitable, how it comes about, what price is paid along the way, and how happy our descendants are about it, will depend very much on choices that are made in the early days—just as large parts of life in the Americas are still shaped (sometimes for better, often for worse) by

---

* Associate Professor of Law, University of Tennessee. Chair, Policy Committee, National Space Society. Member, Vice President's Space Policy Advisory Board, National Space Council, Executive Office of the President. J.D. 1985, Yale Law School; B.A. 1982, University of Tennessee. The opinions here are my own and do not necessarily represent those of any organization with which I am affiliated.

1. *Quoted in* NICHOLAS DANILOFF, THE KREMLIN AND THE COSMOS 20 (1972).

2. DANIEL BOORSTIN, THE REPUBLIC OF TECHNOLOGY: REFLECTIONS ON OUR FUTURE COMMUNITY 30 (1978).

the acts and decisions of the *Conquistadores* and their sponsors five centuries ago. We thus have an obligation to do our best, notwithstanding our own ignorance and frailty, to lay a proper foundation. In the pages that follow, I will try to sketch briefly some of the considerations that should guide us in this task. First, I will look at the narrower question of general legal structures and then at the larger question of what kind of space activity overall is suitable.

I. The Meaning of Peace

Peace, at least as I interpret it for the purposes of this Symposium, is both more and less than the absence of war. It is more because peace is a condition in itself, not merely the absence of another condition. The situation that existed between the United States and the former Soviet Union for so many decades could not have been called peaceful, but it did not explode into outright violence; hence the need to coin the term "Cold War" to describe its formally unwarlike, but often deadly, reality. We should aspire to do better than that in the space arena. Peace is less than the absence of war because it is not an absence of conflict. For example, consider the relationship between the United States and France, where conflict is frequent but peace is assured.

Conflict between the United States and France does not lead to war for a variety of reasons, such as shared history and cultural ties. But perhaps the most significant reason is that the two countries know that they would have far more to lose by going to war. The relationship that they have in peace, even if marked by conflict, is too valuable to abandon.

Similarly, we cannot expect to prevent conflicts from developing as humanity expands into outer space. Conflict is a very human thing—and often, when properly channeled, a *good* thing—and it will accompany us wherever we go. But we can try to create a system in which a peaceful relationship is valuable enough to all participants that conflict does not ripen into war. A few examples of such systems follow, together with some thoughts on how to apply the lessons they offer to outer space.

II. Some Analogies

Our record of managing conflict so as to prevent war may not seem all that impressive, given the number of wars that have raged in this century. On the other hand, there are many areas of human activity that have not led to war.

A. Antarctica

Of the Earth's continents, Antarctica is the only one that has never known genuine war. This is probably the reason why some

commentators have suggested applying the international legal regime governing Antarctica to outer space. However, such an application probably would be a bad idea.

Under existing international law, Antarctica is off-limits to international sovereignty (though preexisting claims are not extinguished, merely rendered dormant).[3] In this respect, Antarctica's legal status resembles that already obtained for outer space under the 1967 Outer Space Treaty (with the exception that there are no dormant preexisting claims for outer space).[4] However, existing international law also bans exploitation of Antarctic minerals. This is a good idea with regard to Antarctica, but it is likely a poor idea with regard to outer space.

There are two good reasons for banning the exploitation of Antarctic minerals. The first is environmental: The Antarctic region is the last more-or-less pristine continent, and it plays an extremely significant, and perhaps preeminent, role in the global climate. In light of these factors, changes to the Antarctic environment of the sort that might be brought about by mineral exploitation should be viewed with considerable suspicion, at least until we know considerably more about the likely consequences to the rest of the Earth. The second reason is political: The ban on mineral exploitation makes the preexisting, and now dormant, national claims essentially beside the point. Were nations free to exploit Antarctic minerals, those claims would be reignited, and conflict—perhaps even outright war—might well ensue.

Neither reason applies in the space context. No space activity is likely to have the kind of critical negative impact on global climate systems that industrialization of the Antarctic might have. In fact, the most likely space industrial activities, microgravity materials research in near earth orbit and lunar mineral mining, are unlikely to have any direct physical effect on the Earth's biosphere or climate at all; any indirect effects are likely to be benign, or else actively good. Nor are there any preexisting national claims to the Moon or other celestial bodies that can be reignited by space industrialization. Any such activities will have to be started afresh, under a regime that (as addressed below) can be tailored so as to promote peace rather than undermine it. (This is an advantage of thinking about such issues early on.)

As a result, analogies to the Antarctic legal regime are unlikely to be very fruitful. However, the Antarctic is not the only international common area to which we can look for examples.

---

3. Antarctic Treaty, Dec. 1, 1959, 12 U.S.T. 794, 420 U.N.T.S. 71.

4. Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, Jan. 27, 1967, 18 U.S.T. 2410, 610 U.N.T.S. 205 [hereinafter Outer Space Treaty].

B. The High Seas

The high seas—those parts of the ocean beyond any nation's territory—share with outer space the legal character of *res communis*. That is, they are international common areas that are open to all and free for exploitation by all, but are not subject to national appropriation. All nations may make use of the high seas for navigation and fishing, but they are not free to exclude others from those uses on any kind of long-term basis.

This system has worked fairly well: Although blockades and naval actions have hardly been unknown, relatively few major international conflicts have arisen over the use of the high seas. We seem to have arrived at a system that is benign, in the sense that it does not promote conflict, and stable, in the sense that everyone seems reasonably happy with it. This does not mean an absence of conflict, of course, but rather that disputes over the general principles that should govern use of the high seas have not been major contributors to war.

From this situation we may be able to extract similar principles that will govern outer space with similar happy results.[5] I will discuss this issue in greater detail later.

C. International Airspace

The rules governing international airspace are very similar to those governing the high seas: All nations are free to make use of that airspace and no nation is empowered to exclude others from such use for any extended period of time. One difference is that use of international airspace is more closely governed by international bureaucracy, the International Civil Aviation Organization (ICAO), than is the use of the high seas. There are several justifications for closer governance with airspace: International aviation is newer and less mature technically than maritime activity, the relationship between air carriers and governments has traditionally been closer, and the link between aviation—even civil aviation—and military matters has traditionally been seen as closer. These factors may or may not be as significant with regard to space activity.[6]

III. Managing Outer Space to Promote Peace

In trying to promote peace in the context of outer space activity, I believe that there are two key goals. First, outer space should not

---

5. *See generally* Hamilton DeSaussure, *Maritime and Space Law, Comparisons and Contrasts*, 9 J. Space L. 93 (1981).

6. *See generally* Irvin L. White, Decision-Making for Space: Law and Policies in Air, Sea and Outer Space 179 (1971); Space Activities and Emerging International Law 169 (N. Matte ed., 1984).

become a new arena for old conflicts. Second, it should not become an incubator for new ones. Some years ago, I suggested (along with Robert Merges) that the best way to prevent outer space from becoming a barren battlefield is to promote its commercial development.[7] I still believe this for two reasons: commercial activity in space makes outright conflict there less likely, and it tends to create positive forces for peace that are likely to defuse conflicts on Earth as well.

A. Military Uses of Space

The military uses of outer space are sufficiently significant that total space demilitarization is very unlikely. This means that even if civilian uses of space—whether scientific or commercial—were abandoned, outer space would still remain an important arena for military activity. The question is: what kind of military activity?

Some have actually applauded the idea of outer space as an arena for military conflict. After all, they reason, no one lives there, so military action will be clean—our robots versus their robots, and may the best cybernetic devices win. Though war in space might be common, it would be bloodless war in the ultimate push-button battlefield. This strikes me as a rotten idea; fortunately, it is unlikely to come to pass.

It is a rotten idea because I have no real confidence that military conflicts in outer space will stay confined to outer space. If the stakes are high enough, conflicts will inevitably spill over into other theaters. If the stakes are not high enough, then they are probably not worth the considerable expense of launching space battle fleets, robotic or otherwise.

Military battles in space are unlikely to occur because outer space is already too valuable as a center of commercial activity. Satellite communications alone are a multibillion-dollar-per-year industry, and the value of satellite communications in tying together global industries is far greater than the dollar figure suggests. A major disruption of satellite communications—a near-certain side effect of significant space combat, even among automated devices—would bring global business to a near-standstill in short order, with phenomenal costs. And satellite communications is only one of the many civilian and commercial activities that already take place in outer space, although not necessarily the most valuable activity over the long term.

In short, outer space makes no more sense as an arena for "clean" warfare than do the floors of the world's stock exchanges, and the ultimate consequences of such warfare would be similar. The

---

7. Glenn H. Reynolds & Robert P. Merges, *The Role of Commercial Development in Preventing War in Outer Space*, 25 JURIMETRICS J. 130 (1985).

value of outer space activity for peaceful purposes far outweighs the value of outer space as an arena for combat. This does not, however, rule out all military uses for outer space, nor even suggest that all such uses would be destabilizing or undesirable. For example, many "passive" uses of outer space, such as the stationing of reconnaissance satellites in orbit, probably serve to promote peace and stability on earth by making arms-control agreements easier to enforce and by making surprise attacks more difficult.[8] Similarly, the specialized early-warning and nuclear blast detection satellites probably promote stability by giving their possessors a greater sense of security and by spotting nuclear tests worldwide, however remote or clandestine.[9] Such uses should continue because they are both harmless in themselves and inherently beneficial in their effects. Space, in this sense, should not be "demilitarized," but "pacified."

B. The Promise of Commercial Development

I have already explained why existing commercial activity in outer space makes conflict there less likely. Now I would like to discuss ways in which future commercial activity—the extraction of wealth, probably ultimately along free-enterprise lines—in outer space might make conflict less likely not only in outer space but on Earth. In a very limited way this is already happening: The growth of global telecommunications—largely a product of satellite technology—undoubtedly had a major effect in bringing the Cold War to an end. The potential does not end there, however.

Much more aggressive expansion into space, such as the "space settlement" envisioned by space advocates and recently called for by Congress and the President, is likely to have important stabilizing effects. First, it is likely that aggressive efforts toward civilian activity (both commercial and scientific) in outer space will serve as a distraction of sorts, directing energies (and industries) that might otherwise fuel militarism into more productive and less threatening directions. I cannot speak scientifically of this likelihood because I am talking about politics, not technology. But many writers have spoken of the formation of relationships among government officials, contractors, and the military as "iron triangles" supporting the growth of the military-industrial complex and furthering tendencies toward war. Such "iron triangles" were inevitable in the Cold War

---

8. *See* Robert P. Merges & Glenn H. Reynolds, *News Media Satellites and the First Amendment: A Case Study in the Treatment of New Technologies*, 3 HIGH TECH. L.J. 1 (1988).

9. *See* PAUL STARES, SPACE AND NATIONAL SECURITY (1987); WILLIAM BURROWS, DEEP BLACK: SPACE ESPIONAGE AND NATIONAL SECURITY (1986); Glenn H. Reynolds, *National Security on the High Frontier*, 2 HIGH TECH. L.J. 281 (1987).

because all major government programs (military or otherwise) tend to develop self-reinforcing constituency groups over time: groups that support expansion (or at least continuation) of efforts and that tend to find new reasons for existing after the old reasons have expired.

Rather than bemoaning this characteristic of politics—which is likely rooted in inherent human characteristics[10]—I suggest that we consider putting it to use. While not all members of the military-industrial complex will accept the substitution of civilian space-exploration spending for military spending, many will (especially among the contractors). The result will be a fracturing and diminution of the iron triangles that support warlike efforts, and the creation instead of what analyst Daniel Deudney has called "iron triangles for peace."[11]

There is precedent for this approach: According to many who were present at the time, John F. Kennedy very deliberately used the Apollo program as a way of distracting attention from the Cold War and of causing the death of military space programs then under consideration (such as SAINT, a satellite interceptor, and BAMBI, an antimissile program).[12] The plan succeeded not only in terms of inside-the-beltway politics, but also in terms of general social attitudes. As Freeman Dyson recounts:

> Apollo was to be what William James had called for long ago, a moral equivalent of war. The idea was to escape from the stuckness of Soviet-American political quarrels by beating the Russians in a bloodless technological competition instead of beating them in battle. The idea was a good one, and up to a point it worked. It stopped working when the symbolic battle of Apollo was displaced from the focus of public attention by the real battle of Vietnam. Unfortunately, nobody since 1961 has repeated Kennedy's tactic of deliberately committing a country to a daring nonmilitary enterprise as a substitute for the excitements of war. It is a tactic which we could profitably use again.[13]

Of course, the problem now is to not create a distraction from Cold War tensions; thankfully, we appear to be past that need. Instead, the problem is to lay a foundation that will promote peace in the

---

10. I am reminded of the scene in Mel Brooks' *Blazing Saddles* in which the Governor brings his cronies together for a strategy session and exclaims, "Gentlemen, we've got to protect our phony-baloney jobs!"

11. Daniel Deudney, *Forging Missiles into Spaceships*, 2 WORLD POL'Y J. 271 (1985).

12. For more on this, see Glenn H. Reynolds, *National Security on the High Frontier*, 2 HIGH TECH. L.J. 281 (1988); Glenn H. Reynolds, *Structuring Development in Outer Space: Problems of How and Why*, 19 LAW & POL'Y INT'L BUS. 433 (1987).

13. FREEMAN DYSON, WEAPONS AND HOPE 219 (1984).

future and provide for whatever new tensions that develop—and develop they inevitably will.

Instead of competing with the Russians, we might consider cooperating with them. Not only would such a venture be a lasting monument to cooperation and good will, it would also make much sense from a technological standpoint. The former Soviet Union developed many technologies (such as cheap expendable launch vehicles, and long-term space habitats) that the West needs; the West, on the other hand, has many fields of expertise that mesh well with the Soviet-developed technologies, and the West also has money. Cooperative ventures would also give many scientists and aerospace engineers—both Eastern and Western—something to do now that Cold War defense budgets are winding down.

This would not merely be a jobs program, but a means of keeping painfully and expensively developed human, intellectual, and technological capital from being squandered. It would also have the advantage—particularly with regard to members of the former Soviet space program—of keeping laid-off workers from peddling their expertise to the next most likely clients, third-world nations trying to develop ballistic missile programs. Rocket-making expertise is (alas) readily transferable to missile-making, and a few good foreign technical experts can move a program along in a hurry, as the United States' experience with Wernher von Braun and his compatriots made very clear.[14] It is far better—and far cheaper for the West over the long run—for those experts to put their skills to use in peaceful ways. Such efforts will make profitable use of capital built up during the Cold War to lay the foundation for later efforts by private organizations and enterprises.

Over the longer term, aggressive space efforts may play a different kind of role in promoting world peace. They may do so by providing wealth. This is not a new dream, but an old one, as the quotation from Tsiolkovsky at the beginning of this Article demonstrates. Earthbound economies are limited to the resources available on or near the surface of the Earth. Alas, as human populations grow, these resources do not. If all nations on the Earth are to enjoy the kind of wealth that Westerners enjoy today, earthly resources are unlikely to suffice, or if they do, they will be recoverable only at ruinous environmental costs, which are themselves likely to breed nasty confrontations. But, as Gregg Easterbrook has pointed out, we

---

14. For more on this see Karp, *The Commercialization of Space Technology and the Spread of Ballistic Missiles in* INTERNATIONAL SPACE POLICY: LEGAL, ECONOMIC AND STRATEGIC OPTIONS FOR THE TWENTIETH CENTURY AND BEYOND 203 (Daniel Papp & John McIntyre, eds. 1987); Jack McCall, *"The Inexorable Advance of Technology?" United States and Multilateral Efforts to Curb Ballistic Missile Proliferation and its Consequences*, 32 JURIMETRICS J. 387 (1992).

are not limited to the resources of the Earth. Easterbrook captures this nicely, writing in a recent issue of *The New Republic:*

> Today we regard the living shroud around Earth and speak as though that were nature. Look up over your head some evening. Earth is one-zillionth of a zillionth of nature. Environmentalists admonish that the only way to understand the environment is to remember that everything is part of everything else. Then they speak as though Earth's ecosphere were a closed system, a be-all and end-all. It is but a tiny part of a vastly greater natural scheme.[15]

If space development is pursued, within a relatively short time (a few hundred years at most, probably much sooner) we will have access to all the material and energy wealth of the solar system. Access to that much wealth (along with the banishing of many dirty processes to locations outside the Earth's biosphere) will allow us to lift the world's poorest nations out of poverty without destroying the Earth's environment and thus allow us to avoid nasty conflicts that would otherwise be bred by scarcity.

Such a future certainly seems preferable to the alternatives: (1) Western nations telling the less developed countries that "the world can afford for us to live well, but not for you to as well"; (2) wars over resources and environmental damage; (3) a decline in living standards all over the globe; or (4) apocalyptic damage to the global environment. Obviously, aggressive efforts to exploit space resources are no alternative to the necessary short-term remedies of conservation and increased attention to environmental hygiene. However, just as obviously, those short-term methods hold no promise of long-term salvation; however much we conserve, we will eventually run out. And the various sacrifices that will be necessary to implement such measures (in both the wealthy and the poor nations) will be easier to bear if the general perception is that the pie is growing rather than shrinking or remaining stagnant. Such a perception should reduce the tendency toward violent conflict.

IV. MAKING IT WORK

Of course, development in outer space must be properly structured if it is to have these productive effects. If improperly structured, development of outer space resources could breed violent conflict, rather than prevent it. If human history is any guide, after all, anything worth having is worth fighting over, and efforts at development would make space resources worth having. Furthermore,

---

15. Gregg Easterbrook, *Everything You Know About the Environment is Wrong*, THE NEW REPUBLIC, April 30, 1990, at 14, 27. *See also* Sen. Albert Gore, Jr., *Outer Space, The Global Environment, and International Law: Into the Next Century*, 57 TENN. L. REV. 329 (1990).

space resources would not make poor countries wealthier, or more confident that their share of the pie would ultimately grow, if the benefits were hogged by a few wealthy nations.

There has already been some effort to deal with this problem, though in a way that I regard as misguided, through the 1979 Moon Treaty.[16] That treaty provides (as the earlier and more successful 1967 Outer Space Treaty does not) that private property rights in space resources are forbidden, and that such resources may be exploited only by an international authority that apportions its profits in such a way as to give less developed countries a substantial rake-off regardless of whether they play any role in wealth creation. Under the Moon Treaty, the international authority presumably would ensure that poorer countries received some wealth, and its role as the only game in town presumably would prevent efforts at lunar land-grabs and such. Unfortunately, the Moon Treaty's close attention to pie-dividing would likely discourage those with the necessary capabilities from doing much to develop the resources in question to begin with.

Fortunately, no major space power has joined the Moon Treaty, and it is unlikely to play a significant role in governing space resource development. This still leaves us with the question of what to do, however. This is not the place to lay that out in detail,[17] but here are some thoughts.

First, no international authority should have exclusive rights to extract space resources. Such an exclusive role would contravene the 1967 Outer Space Treaty's provision that "Outer space, including the moon and other celestial bodies, shall be free for exploration and use by all states, without discrimination of any kind, on a basis of equality and in accordance with international law, and there shall be free access to all areas of celestial bodies."[18] An exclusive role for an international body would also eliminate the spur of competition, which likely would slow down the rate of progress and hence the rate at which benefits returned to Earth. One lesson of the latter part of this century, after all, is that large bureaucracies often become flabby and moribund, and that state enterprises often lack the will or ability to flourish.

Second, there should be broad participation by less developed countries, but on an at-risk basis. They should be real participants, not merely silent beneficiaries of any profits that might (or might

16. Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, U.N. Doc. A/AC 105/L.113/Add.4 (1979).

17. For more on this topic see GLENN H. REYNOLDS & ROBERT P. MERGES, OUTER SPACE: PROBLEMS OF LAW AND POLICY 102-166 (1989); Glenn H. Reynolds, *The International Law of Outer Space: Into the Twenty-First Century*, 25 VAND. J. TRANSNAT'L L. 225 (1992).

18. Outer Space Treaty, *supra* note 4, art. I.

not) emerge. This participation should be made feasible by low-cost financing, by allowing the less developed countries to make in-kind contributions (of, for example, launch sites near the equator), or by other mechanisms that would lower the barriers to entry without making the countries less than full partners in the venture. Real participation would give everyone an incentive to see the venture succeed. Exclusion—or on the other hand, the mere grant of a rake-off of possible profits, as provided in the Moon Treaty—merely would breed resentment on one side or another.

Third, to encourage development, there should be land grants or other similar grants of interest in space resources to serve as an incentive for development. The use of land grants to railroads was instrumental in opening the American West; properly structured, such grants could dramatically increase the pace at which space resources are developed, and they could draw private capital and expertise into ventures that would otherwise be undertaken only by government agencies or not at all. Such grants could be administered by a variety of entities, from the United Nations or another multinational body to the United States government (which could not claim sovereignty over space resources itself, but which could recognize claims by United States citizens, a position analogous to the government's existing position with regard to the deep seabed). Such grants should be contingent on development within a reasonable time (to ensure results and to prevent undue speculation), but they should be freely alienable, so as to promote capital formation.

Fourth, there should be adequate dispute-resolution mechanisms available, so that disputes (for example, over boundaries) do not ripen into violent conflict. This will be a tough problem because international dispute-resolution mechanisms are currently poorly developed. But it should not be an insoluble one, given the relatively narrow range of likely disputes and their probable technical nature.

This Article is merely the roughest sketch of what a successful space-development regime should look like. However, all great works of art are preceded by many sketches. Over the next decade or so, many lawyers, policymakers, and academics will be outlining the course of humanity's expansion into space. I hope that others will sharpen their pencils and join in the project.

Protection of Biological and Cultural Diversity: Emerging Recognition of Local Community Rights in Ecosystems Under International Environmental Law

LEE P. BRECKENRIDGE*

INTRODUCTION

The United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in June 1992 forged a new consensus on international environmental policies to protect the world's biological diversity and its most fragile ecosystems. The documents produced by UNCED (the Rio Declaration,[1] Agenda 21,[2] and the Forest Principles[3]) and the two conventions opened for signature at the conference (the Convention on Biological Diversity[4] and the Convention on Climate Change[5]) all take important steps toward the

---

* Associate Professor of Law, Northeastern University School of Law. J.D. 1976, Harvard Law School; B.A. 1973, Yale University.

1. Adoption of Agreements on Environment and Development: The Rio Declaration on Environment and Development, U.N. Doc. A/CONF.151/5/Rev.1 (1992) [hereinafter Rio Declaration] (on file with *Tennessee Law Review*).

2. Agenda 21 (prov. ed. July 10, 1992), *available in* EcoNet, en.unced.docum conference, File Nos. 409-48 [hereinafter Agenda 21] (on file with *Tennessee Law Review*). The various chapters of Agenda 21, adopted on June 14, 1992, were issued in an advanced version, subject to the following note: "This document will be further edited, translated into the official languages, and published by the United Nations for the General Assembly this autumn." *Id.* ch. 1. The final publication of Agenda 21 was consequently unavailable for this article. For the preceding draft of Agenda 21, see Adoption of Agreements on Environment and Development: Agenda 21: Note by the Secretary-General of the Conference, A/CONF.151/4 (Parts I-IV) (1992) (transmitting the draft text of Agenda 21 as approved by the Preparatory Committee for UNCED at its fourth session).

3. Adoption of Agreements on Environment and Development: Non-legally binding authoritative statement of principles for a global consensus on the management, conservation and sustainable development of all types of forests, U.N. Doc. A/CONF.151/6/Rev.1 (1992) [hereinafter Forest Principles] (on file with *Tennessee Law Review*).

4. Convention on Biological Diversity, *opened for signature* June 5, 1992 (on file with *Tennessee Law Review*).

5. United Nations Framework Convention on Climate Change, *opened for signature* June 3, 1992 [hereinafter Convention on Climate Change] (on file with *Tennessee Law Review*). The text of the Convention on Climate Change is set forth in Annex I to the Report of the Intergovernmental Negotiating Committee for a Framework Convention on Climate Change on the Work of the Second Part of its Fifth Session, Held at New York from 30 April to 9 May 1992, U.N. Doc. A/AC.237/18 (Part II)/Add.1 (1992). The Convention on Biological Diversity and the Convention on Climate Change were negotiated outside the UNCED Preparatory Committee meetings, through separate international negotiating committees.

formulation of international standards governing the use and management of living resources to preserve their diversity and renewability into the future.

The articulation of international environmental requirements is accompanied, strikingly, by a new recognition of local communities' roles in protecting biological diversity and ecosystem viability. "Grassroots" empowerment has become a centerpiece of the environmental agenda. Throughout the UNCED documents,[6] a mandate for decentralization goes hand-in-hand with the centralization expressed in new international environmental norms and institutional mechanisms. International environmental law is emerging as a new source of authority for pluralism, and protection of biological diversity has become inextricably linked to protection of cultural diversity.

The convergence of local community goals and international environmental aims was illustrated by the gathering of representatives of non-governmental organizations that took place in Rio de Janeiro during the official events of UNCED. The "Global Forum" united environmental activists urging recognition of global ecological interdependence, and human rights activists urging protection for rights of indigenous peoples and other marginalized communities to their traditional lands.[7] The recent alliances among such groups have embodied a linkage of environmental and human rights goals that also converged in the official negotiations of government representatives.

This Article examines the alliance between the themes of biological diversity and cultural diversity[8] in the provisions of the UNCED

---

6. The Convention on Biological Diversity, the Convention on Climate Change, the Rio Declaration, Agenda 21, and the Forest Principles are referred to hereinafter, collectively, as "the UNCED documents."

7. Many of these groups were also accredited as observers in the governments' official proceedings. A vivid impression of how UNCED brought human rights and environmental activists together in an unprecedented way can be gained simply by reviewing the list of organizations recommended for accreditation in the course of the UNCED preparations. *See* U.N. Doc. A/CONF.151/PC/L.28/Adds. 1-14 (1992). For commentary on the implications of the "partnerships" forged in preparations for the Global Forum, see W. H. Lindner, When Asking the Questions is Part of the Answer, Network '92, May 1992, at 3 (newsletter published by the Centre for Our Common Future) (on file with *Tennessee Law Review*). The increasingly far-reaching efforts of United States environmental organizations to build alliances with human rights and environmental activists in other countries are recounted in David A. Wirth, *Legitimacy, Accountability, and Partnership: A Model for Advocacy on Third World Environmental Issues*, 100 YALE L.J. 2645 (1991) [hereinafter Wirth].

8. This Article does not attempt to address the more general question of how international environmental law has become linked to international human rights law and rights to economic development. The focus of this article is narrower, emphasizing ways in which emerging international environmental law to protect biological diversity has become explicitly allied with efforts to protect the distinctive

documents relating to management of biological resources. Part I provides a preface to the discussion, examining two lines of legal analysis that became influential at UNCED. This part explores separately an "environmental" and a "human rights" perspective on rights to biological resources. Emerging principles of international environmental law have advanced notions of global trusteeship in the management of biological resources, while international human rights documents have supported recognition of local rights of access and management authority. These two approaches, while distinct, have been mutually reinforcing in their challenges to existing concepts of sovereignty and private property.

Part II discusses the alliance and synergism of these two perspectives at UNCED, showing how a central international environmental standard of "sustainability" has become inextricably linked to the empowerment of local communities and especially to the formulation of rights for indigenous peoples and other marginalized communities.

Part III concludes that the alliance of global environmental goals and local community rights has important implications for the management and use of living resources. To illustrate some of the implications, Part III examines briefly some models of resource management that might be cited as exemplifying the linkage advanced in the UNCED documents. Part III suggests, however, that there are tensions in an alliance that might, at first glance, appear to blend harmoniously the aspirations of local communities with the aims of international environmental policy. The two perspectives are closely allied in their shared repudiation of environmental and human havoc wrought in the name of state sovereignty or state-created definitions of private property. Yet the meaning of "autonomy" and "self-

---

identities of local communities, particularly indigenous peoples and other "traditional" communities marginalized by dominant societies. For commentary on the broader issues, see Russel Lawrence Barsh, *The Right to Development as a Human Right: Results of the Global Consultation*, 13 HUM. RTS. Q. 322 (1991); *Environment, Economic Development and Human Rights: A Triangular Relationship?*, 82 PROC. AM. SOC'Y INT'L L. 40 (1988) [hereinafter *Environment, Economic Development and Human Rights*]; Review of Further Developments in Fields with Which the Sub-Commission Has Been Concerned: Human Rights and the Environment: Progress Report prepared by Mrs. Fatma Zohra Ksentini, Special Rapporteur, pursuant to Sub-Commission resolution 1991/24, U.N. Doc. E/CN.4/Sub. 2/1992/7 (1992); Sierra Club Legal Defense Fund, Human Rights and the Environment: The Legal Basis for a Human Right to the Environment: Report to the United Nations Sub-Commission on the Prevention of Discrimination and the Protection of Minorities (April 1992) [hereinafter Sierra Club Legal Defense Fund] (on file with *Tennessee Law Review);* Review of Further Developments in Fields with Which the Sub-Commission Has Been Concerned: Human Rights and the Environment: Preliminary report prepared by Mrs. Fatma Zohra Ksentini, Special Rapporteur, pursuant to Sub-Commission resolutions 1990/7 and 1990/27, U.N. Doc. E/CN.4/Sub.2/1991/8 (1991) [hereinafter Ksentini Preliminary Report].

determination" of local communities in a world of ecological interdependence, international responsibility, and the search for global community, remains to be elaborated. The UNCED documents reflect the beginnings of a discussion about the autonomy and the interdependence of the human place in nature, rather than a definitive solution to questions of jurisdiction and ownership.

I. AUTHORITY OVER BIOLOGICAL RESOURCES: AN "ENVIRONMENTAL" AND A "HUMAN RIGHTS" PERSPECTIVE ON BIOLOGICAL DIVERSITY AND ECOSYSTEM VIABILITY

The conjunction of biological diversity and cultural diversity themes in the UNCED documents might be said to draw inspiration from two separate directions. This statement simplifies a complex story; the purpose of this Article, however, is not to provide a full history of events and negotiations leading up to UNCED, but rather to highlight a convergence of themes in the UNCED documents that is significant for the future development of international environmental law. Defining these two basic approaches, with their distinctions and growing similarities, will help to show more clearly certain implications of the work that was accomplished at UNCED.

On the one hand, we can discern an international "environmental" perspective, reflected in a growing internationalization of environmental law relating to the conservation of living resources. This approach emphasizes worldwide trends in species destruction and ecosystem degradation. It points to expanding scientific information on global ecological interdependence and advocates international measures to deal with losses in biological diversity. It sees biological diversity as a matter of global importance to be managed for the benefit of an international community. For reasons that will be explored below, the "environmental" perspective has also come to see cultural diversity as a global resource integrally related to biological diversity.

On the other hand, we can discern a "human rights" perspective, reflected particularly clearly in international human rights documents dealing with rights of indigenous people and their communities. This approach focuses on the violence, illness, impoverishment, and even death that human communities may suffer when the renewable natural resources of their environment are altered and destroyed. The conservation of biological diversity and the viability of ecosystems is seen as one important aspect of protecting the survival of particular human communities and their distinctive cultures. Local access to resources and exclusion of incompatible uses is viewed as central to preserving the life of the community.

These two approaches begin from different standpoints. They both address rights to biological resources, but they offer two different definitions of the relevant "community." The "environmen-

tal" perspective advances notions of global heritage and trusteeship transcending state boundaries, while the "human rights" perspective insists on local community rights and self-management in the face of competing government and private claims. The "environmental" perspective emphasizes states' obligations to the international community while the "human rights" perspective emphasizes states' duties to local communities. The "environmental" perspective sees cultural diversity as a resource in the preservation of biological diversity; the "human rights" perspective sees it as a matter of community identity.

Despite these differences, the two approaches are convergent. In particular, the two perspectives share a conclusion that states' exercise of sovereign power, and their allocation and enforcement of property rights, have proven in the past to be insufficient and at times directly contrary to the goal of preserving ecosystems and the human communities that depend upon them. The ideas of global environmental rights and local community rights are mutually reinforcing, and the convergence of the two perspectives consequently represents a powerful alliance for change that challenges current government decision-making and seeks solutions through international law.

The following sections examine these two approaches in more detail as a background for considering their emergence in the documents at UNCED.

A. An "Environmental" Perspective: Biological Diversity as a Global Resource

1. *The Meaning and Value of Biological Diversity from an International Standpoint.*—As the scientific understanding of biological diversity has expanded since the 1970s, so too have the policy recommendations for international legal measures addressing the conservation of biological diversity as an environmental issue of global importance.

Biological diversity, or biodiversity, is an umbrella term used to describe the variety and variability of ecosystems, species and genes in nature. The study of biological diversity encompasses not only the numbers and variations of genes within species, and of species within regions, but also the associations of species within communities, and their interactions with one another and with the physical environment in ecosystems.[9]

---

9. JEFFREY A. MCNEELY ET AL., CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY 17-18 (1990) [hereinafter CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY]; WORLD RESOURCES INSTITUTE ET AL., GLOBAL BIODIVERSITY STRATEGY 2 (1992) [hereinafter GLOBAL BIODIVERSITY STRATEGY]. Professor E. O. Wilson provides a vivid account of the emergence of diverse forms of life on earth in his most recent book, EDWARD O. WILSON, THE DIVERSITY OF LIFE (1992) [hereinafter THE DIVERSITY OF LIFE].

The term "biological resources" is often used to describe genes, species and ecosystems, and their ecological complexes, from the standpoint of their value to human beings.[10] People depend on biological resources as sources of food, medicines and industrial products. The existence of wide genetic variety is critical to innovation and adaptability in agriculture and biotechnology. Biological resources provide the basis for tourism, recreation, and aesthetic enjoyment. All of these uses are of obvious importance to international as well as national economies.[11]

More broadly, biological diversity provides a wide range of ecosystem functions or "environmental services." Forests and wetlands, for example, serve to fix solar energy through photosynthesis, maintain water cycles, produce and protect soil, store and recycle nutrients, provide nursery and breeding grounds, and absorb and break down pollutants. Forest ecosystems play a global role as carbon sinks that draw carbon dioxide out of the atmosphere, reducing the concentration of gases that would lead to warming of the atmosphere, as well as a regional role in regulating climate.[12]

The total number of species on earth is not known, and even the interrelationships among the species that have been identified are not well understood. Consequently the actual and potential benefits of biological diversity, both direct and indirect, to human beings cannot

---

10. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 18. The emphasis on biological "resources" in the present discussion reflects the focus that prevailed at UNCED. The term "resources," with its anthropocentric and utilitarian connotations, is used throughout the UNCED documents. Although many have argued that people and states cannot truly own or control nature (*see generally*, RODERICK FRAZIER NASH, THE RIGHTS OF NATURE (1989); RODERICK FRAZIER NASH, WILDERNESS AND THE AMERICAN MIND (1967) (discussing arguments advanced by activists and writers in the United States)), the UNCED documents place human beings at the center of their focus, while defining broadly the "goods and services" provided to human beings by the rest of the natural world in sustaining conditions suitable for human existence. *See, e.g.*, Rio Declaration, *supra* note 1, Principle 1 ("Human beings are at the centre of concerns for sustainable development"); Agenda 21, *supra* note 2, para. 15.2 ("goods and services"); Forest Principles, *supra* note 3, Principles/Elements para. 2(b) ("forest products and services").

11. THE DIVERSITY OF LIFE, *supra* note 9, at 281-305; CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 28-31, 34; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 2-5; WALTER V. REID & KENTON R. MILLER, KEEPING OPTIONS ALIVE 22-30 (1989) [hereinafter KEEPING OPTIONS ALIVE]. On the centrality of biological diversity to agricultural innovation, see also CALESTOUS JUMA, BIOLOGICAL DIVERSITY AND INNOVATION: CONSERVING AND UTILIZING GENETIC RESOURCES IN KENYA (1989).

12. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 32-33; KEEPING OPTIONS ALIVE, *supra* note 11, at 4-8; THE DIVERSITY OF LIFE, *supra* note 9, at 305-09. For discussion of additional values of biological diversity, see *Paul R. Ehrlich & Anne H. Ehrlich, The Value of Biodiversity*, XXI Ambio 219 (1992); BRYAN G. NORTON, WHY PRESERVE NATURAL VARIETY? (1987); CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 33-35.

be fully known. However, enough is known to see clearly that conserving biological diversity is central to preserving the "life-support systems" of the earth and that the management of biological resources has long-term and global implications, not just for present generations of human beings, but for future generations as well.[13]

Biological diversity is being rapidly destroyed as a result of human practices. Habitat destruction and fragmentation, over-exploitation of plant and animal species, introduction of foreign species, pollution, and industrial agriculture and forestry practices are primary causes of species extinctions.[14] One recent estimate of species extinction concluded that five to fifteen percent of all species would be lost between 1990 and 2020.[15] Extinction rates are particularly high in tropical forests, but biological diversity is declining throughout the world. Given the interdependence of species and ecosystems, the "cascade" effects of species loss may be dramatic.[16]

These, in brief, are some of the central reasons for a growing international concern about the ways that states exercise sovereign authority over the biological resources within territorial jurisdiction. International environmental norms have been emerging slowly in response to the scientific consensus.

2. *The Emergence of International Standards for Conservation of Biological Diversity.*—The principle of national sovereignty over natural resources has stood as an obstacle to formulating general international standards to preserve biological diversity. Most of the world's areas of greatest biological diversity are located on land, and

---

13. THE WORLD CONSERVATION UNION ET AL., CARING FOR THE EARTH: A STRATEGY FOR SUSTAINABLE LIVING 27-28 (1991) [hereinafter CARING FOR THE EARTH] ("Life-support systems are the ecological processes that shape climate, cleanse air and water, regulate water flow, recycle essential elements, create and regenerate soil, and keep the planet fit for life. . . . Plants and animals, evolving over hundreds of millions of years, have made the planet fit for the forms of life we know today."). *See also*, CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 18; THE DIVERSITY OF LIFE, *supra* note 9, at 15, 347.

14. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 37-45; THE DIVERSITY OF LIFE, *supra* note 9, at 253-72; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 7-12.

15. KEEPING OPTIONS ALIVE, *supra* note 11, at 37-38. Estimates put the number of species living on earth at 10 million or more, although only 1.4 million have been identified. KEEPING OPTIONS ALIVE, *supra* note 11, at 9; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 7-9; THE DIVERSITY OF LIFE, *supra* note 9, at 133.

16. KEEPING OPTIONS ALIVE, *supra* note 11, at 22-30; E. O. Wilson, *The Current State of Biological Diversity*, *in* BIODIVERSITY 3-18 (E. O. Wilson ed., 1988) [hereinafter BIODIVERSITY]. "This utter dependence of organisms on appropriate environments is what makes ecologists so certain that today's trends of habitat destruction and modification . . . are an infallible recipe for biological impoverishment." Paul R. Ehrlich, *The Loss of Diversity: Causes and Consequences*, *in* BIODIVERSITY, *supra*, at 22.

in freshwater and coastal marine environments, within state boundaries. Principle 21 of the Stockholm Declaration[17] adopted by the United Nations Conference on the Human Environment in 1972, explicitly confirmed states' sovereign rights to exploit their own natural resources, subject to an obligation not to cause extraterritorial environmental damage.[18] States have continued to resist international measures that might intrude on the conduct of mining, logging, and other forms of resource exploitation deemed to be in the national interest.

Despite the purportedly sharp boundaries between internal matters subject to sovereign authority and external impacts subject to international control, however, the lines of distinction have been eroding as the international implications of biological diversity have been better understood. A number of international agreements have addressed particular aspects of conserving biological diversity. The Convention Concerning the Protection of the World Cultural and Natural Heritage[19] establishes a program for designating unique natural areas and cultural sites on a World Heritage List;[20] it recognizes the obligation of states to protect such areas,[21] and of the international community to help pay for them.[22] The Convention on Wetlands of International Importance[23] calls for the wise use of wetlands and the designation of wetlands of international importance, while the Convention on Conservation of Migratory Species of Wild Animals[24] requires protection of endangered migratory species and

---

17. Declaration of the United Nations Conference on the Human Environment [hereinafter Stockholm Declaration] *in* Report of the United Nations Conference on the Human Environment, at 5, U.N. Doc. A/CONF.48/14/Rev. 1 (1973) [hereinafter Stockholm Report].

18. The principle has been reiterated, *e.g.*, in Permanent Sovereignty Over Natural Resources, G.A. Res. 3171, 28 U.N. GAOR Supp. No. 30, at 52, U.N. Doc. A/9030 (1973), and the Charter of Economic Rights and Duties of States, G.A. Res. 3281, U.N. GAOR 29th Sess., Supp. No. 31, at 50, U.N. Doc. A/9631 (1975), *reprinted in* 14 I.L.M. 251 (1975). For observations about the principle of sovereignty over natural resources and the obstacles it poses to the emergence of new requirements in international environmental law, see Jonathan Green & Philippe Sands, *Establishing an International System for Trading Pollution Rights*, 15 Int'l Env. Rep. (BNA) No. 3 at 80, 83-84 (Feb. 12, 1992).

19. Convention Concerning the Protection of the World Cultural and Natural Heritage, Nov. 23, 1972, 27 U.S.T. 37 [hereinafter World Heritage Convention].

20. *Id.* art. 11, 27 U.S.T. at 43.

21. *Id.* Preamble, 27 U.S.T. at 40.

22. *Id.* art. 15, 27 U.S.T. at 45.

23. Convention on Wetlands of International Importance Especially as Waterfowl Habitat (the Ramsar Convention), Feb. 2, 1971, 996 U.N.T.S. 245, T.I.A.S. No. 11,084.

24. Convention on the Conservation of Migratory Species of Wild Animals, June 23, 1979, 19 I.L.M. 15.

calls for further international conservation agreements governing vulnerable species. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)[25] regulates trade in specific species with the goal of ensuring their survival. Numerous other regional measures[26] and agreements addressing specific species[27] have been concluded.[28] The United Nations General Assembly broadly expressed its support for conservation of genes, species, and ecosystems in the World Charter for Nature in 1982,[29] reaffirming concerns about maintaining the biological resources of the earth that were stated a decade before in the Stockholm Declaration.[30]

The patchwork of international measures developed prior to UNCED has failed to protect effectively many types of species, habitats, and ecosystem functions. Collectively, however, these documents have reflected an emerging international consensus about the responsibilities of states to maintain biological diversity and the viability of ecosystems.

Commentators have suggested that a framework has been evolving under customary international law to require states, as trustees, to protect global environmental resources in general; these are resources "located within the territory of one country but broadly enjoyed, and arguably needed, by the world community as a whole."[31] Put another way, the international community is seen as holding rights to natural resources within sovereign territory:

> A global environmental right arises in connection with a global environmental resource. It refers to the right of all states to expect that the resource will be protected by the state in which it is found. States are trustees, responsible for the preservation of species within their territories. That obligation runs to the international community

---

25. Convention on International Trade in Endangered Species of Wild Fauna and Flora, Mar. 3, 1973, 27 U.S.T. 1087, 993 U.N.T.S. 243, *reprinted in* 12 I.L.M. 1085.

26. *E.g.*, Convention on Nature Protection and Wildlife Preservation in the Western Hemisphere, Oct. 12, 1940, 56 Stat. 1354, 161 U.N.T.S. 193; African Convention on the Conservation of Nature and Natural Resources, Sept. 15, 1968, 1001 U.N.T.S. 3.

27. *E.g.*, International Convention for the Regulation of Whaling, Dec. 2, 1946, 62 Stat. 1716, 161 U.N.T.S. 72; International Convention for the Protection of Birds, Oct. 18, 1950, 638 U.N.T.S. 185.

28. For commentary on existing international measures, see CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 137-39; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 62-66.

29. World Charter for Nature, Principles 1-4, G.A. Res. 37/7, 37 U.N. GAOR, 37th Sess., Supp. No. 51, at 17, U.N. Doc. A/37/51 (1982), *reprinted in* 22 I.L.M. 455 (1983).

30. Stockholm Declaration, *supra* note 17, Principles 2, 3, 4, *in* Stockholm Report, *supra* note 17, at 4.

31. Michael J. Glennon, *Has International Law Failed the Elephant?*, 84 AM. J. INT'L L. 1, 34 (1990) [hereinafter Glennon].

> as a whole: any state should be regarded as suffering legally cognizable injury when that obligation is breached by another state.[32]

Under this theory, the biological diversity within a state is part of a common heritage that is held in trust.[33] The concept of trusteeship has been further extended to encompass obligations to future generations as well as to current inhabitants of the planet. According to this view, not only do states owe duties to each other, but the entire world community bears a responsibility to ensure the renewability of global resources for future generations.[34]

When UNCED took up the task of articulating international standards to govern the management of biological resources, it was building in part on this emerging international consensus regarding state responsibilities and global rights. A vision of global ecological interdependence and international community lies at the heart of this consensus. States continue to hold "sovereign" authority over natural resources, but the power to exploit is increasingly constrained by duties to a wider community requiring preservation of biological diversity and maintenance of ecosystem functions.

3. *Cultural Diversity as a Global Resource Related to Biological Diversity.*—If ecosystems are global environmental resources, then the roles that human inhabitants play in managing and altering those

---

32. *Id. See also*, David D. Caron, *The Law of the Environment: A Symbolic Step of Modest Value*, 14 YALE J. INT'L L. 528, 529 (1989); Ved P. Nanda & William K. Ris, Jr., *The Public Trust Doctrine: Viable Approach to International Environmental Protection*, 5 ECOLOGY L.Q. 291, 294 (1976) [hereinafter Nanda & Ris]. Compare, however, the view expressed in WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT, OUR COMMON FUTURE 162 (1987) [hereinafter OUR COMMON FUTURE]: "Collective responsibility for the common heritage would not mean collective international rights to particular resources within nations. This approach need not interfere with concepts of national sovereignty. But it would mean that individual nations would no longer be left to rely on their own isolated efforts to protect species within their borders."

33. EDITH BROWN WEISS, IN FAIRNESS TO FUTURE GENERATIONS: INTERNATIONAL LAW, COMMON PATRIMONY, AND INTERGENERATIONAL EQUITY (1989) [hereinafter IN FAIRNESS TO FUTURE GENERATIONS]; Edith Brown Weiss, *The Planetary Trust: Conservation and Intergenerational Equity*, 11 ECOLOGY L.Q. 495 (1984) [hereinafter *The Planetary Trust*]; *Developments in the Law—International Environmental Law*, 104 HARV. L. REV. 1521, 1533-35 (1991) [hereinafter *Developments in the Law*]. For a discussion of the justifications for recognizing public property rights in ecological systems under United States law and a review of recent legal literature, see Alison Rieser, *Ecological Preservation as a Public Property Right: An Emerging Doctrine in Search of a Theory*, 15 HARV. ENVTL. L. REV. 393 (1991) [hereinafter Rieser]. Similar justifications underlie the shift toward recognition of trusteeship responsibilities owed to an international community. *See* Nanda & Ris, *supra* note 32, at 303-04.

34. IN FAIRNESS TO FUTURE GENERATIONS, *supra* note 33, at 47; *The Planetary Trust*, *supra* note 33, at 498; *Developments in the Law*, *supra* note 33, at 1539-42. *See also*, Philippe J. Sands, *The Environment, Community and International Law*, 30 HARV. INT'L L.J. 393 (1989).

resources become an inevitable topic of interest. Advocates of new international environmental norms have argued that the preservation of biological diversity must logically include the preservation of diverse human ecological niches. This viewpoint is well summarized in a statement of the *Global Biodiversity Strategy*, issued by the World Resources Institute, The World Conservation Union and the United Nations Environment Programme in early 1992, just before UNCED: "Human *cultural diversity* could also be considered part of biodiversity. Like genetic or species diversity, some attributes of human cultures (say, nomadism or shifting cultivation) represent 'solutions' to the problems of survival in particular environments. And like other aspects of biodiversity, cultural diversity helps people adapt to changing conditions."[35]

Cultural traits affecting a community's use and conservation of biological resources are objects of ecological study in this analysis. Cultural diversity is viewed as an aspect of biodiversity: it is a valuable resource to the global community that needs to be conserved and supported within the emerging international environmental framework for protection of biological diversity.[36]

The acknowledgement of a role for local communities in the protection of biological diversity was something of a revolution in international environmental policy. Efforts in the past to protect biological diversity in parks and preserves have often involved the exclusion of all human habitation through government fiat. This earlier approach saw national governments and their agencies as the chief actors in nature conservation efforts, and the members of local communities as a threat to be excluded.[37]

The more recent "environmental" perspective, however, treats local human inhabitants as unavoidable participants in ecosystems.[38]

---

35. GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 3 (emphasis in original). *See also* Jeffrey A. McNeely, *Conserving Cultural Diversity: How the Variety of Human Experience Can Help Promote Sustainable Forms of Using Natural Resources* (1989), *available in* EcoNet, iucn.news conference, File No. 5 [hereinafter Jeffrey A. McNeely] (on file with *Tennessee Law Review*).

36. World Heritage Convention, *supra* note 19, reflects a recognition of the global values of cultural as well as natural heritage, although it does not treat cultural diversity as an integral aspect of biological diversity. For arguments that the two themes should be viewed as closely related, see IN FAIRNESS TO FUTURE GENERATIONS, *supra* note 33, at 257-78; *The Planetary Trust*, *supra* note 33, at 530-31, 559-63.

37. Raymond F. Dasmann, *The Relationship Between Protected Areas and Indigenous Peoples*, *in* NATIONAL PARKS, CONSERVATION, AND DEVELOPMENT, THE ROLE OF PROTECTED AREAS IN SUSTAINING SOCIETY 667-71 (Jeffrey A. McNeely & Kenton R. Miller eds., 1984) [hereinafter NATIONAL PARKS, CONSERVATION, AND DEVELOPMENT]; CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 49; THE DIVERSITY OF LIFE, *supra* note 9, at 282.

38. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 51; THE DIVERSITY OF LIFE, *supra* note 9, at 283.

It particularly values the abilities that some human communities have demonstrated in conserving, enhancing and sustainably using the biological diversity of the lands they inhabit. Ecologists now note that biological diversity and ecological stability often coincide with the traditional territories of communities that have successfully relied on the sustained productivity of local renewable resources.[39] These observations draw a distinction between "traditional" societies, with subsistence economies closely tied to local ecosystems, and industrialized societies, which "draw their support not from any one local ecosystem but from the entire capital of the world's living matter."[40] Communities that depend directly on the renewability of biological resources in a given area for their livelihood have a special stake in sustaining and protecting the biological resource base for local uses, while industrialized societies tend to reduce biological diversity through monocultural agriculture and participation in a global exchange economy.[41]

Four characteristics of local communities that survive successfully within biologically diverse and fragile environments are frequently mentioned as making special contributions to the attainment of international objectives:

(a.) *Knowledge.*—Communities that have managed to survive long-term on the basis of local renewable resources in one area have often developed sophisticated knowledge and strategies relating to the ecosystems of which they are a part.[42] The accumulated customary

---

39. For instance, researchers in Central America have found: "There are no other land use models for the tropical rain forest that preserve ecological stability or biological diversity as efficiently as those of the indigenous groups presently encountered there." Brian Houseal et al., *Indigenous Cultures and Protected Areas in Central America*, Cultural Survival Quarterly, March 1985, at 10-19, *quoted in* CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 50. *See also*, Paul A. Olson, *Introduction* to THE STRUGGLE FOR THE LAND 1, 20-21 (Paul A. Olson ed. 1990) [hereinafter THE STRUGGLE FOR THE LAND].

40. Jeffrey A. McNeely, *Biosphere Reserves and Human Ecosystems*, in CONSERVATION, SCIENCE AND SOCIETY: CONTRIBUTIONS TO THE FIRST INTERNATIONAL BIOSPHERE RESERVE CONGRESS, MINSK, BYELORUSSIA/USSR 496 (1984).

41. Richard B. Norgaard, *The Rise of the Global Exchange Economy and the Loss of Biological Diversity*, *in* BIODIVERSITY, *supra* note 16, at 206-11.

42. A report on one island-based community, for instance, noted:

> Environmental knowledge is extremely complex, wide-ranging and logically ordered. For example, more than 80 terms are used to distinguish different tidal and associated sea conditions. . . . Similar information is taught about winds, and the behavior and natural history of fauna and flora and other aspects of the island-sea environment that makes it predictable, useable and understandable. Other island and coastal peoples have similar sophisticated descriptions and understanding of their environments.

Bernard Nietschmann, *Indigenous Island Peoples, Living Resources and Protected Areas* [hereinafter Nietschmann], *in* NATIONAL PARKS, CONSERVATION AND DEVELOPMENT, *supra* note 37, at 340. *See also* ROBERT GOODLAND, TRIBAL PEOPLES AND

knowledge about ecosystems in some communities may far exceed the information that is currently available in scientific literature. Such traditional knowledge is a cultural resource resulting from close observation of natural phenomena and intergenerational transfer of wisdom.[43]

(b.) *Practices.*—The biological diversity of many areas occupied by traditional communities does not stem from a lack of human intervention; rather, it is often fostered by human participation in the workings of the ecosystem. The traditional agricultural and other practices that have evolved through the experience and innovations of generations of inhabitants may be particularly adapted to fostering the productivity and renewability of biological resources in a given locale.[44]

(c.) *Community Organization.*—Many traditional societies have developed complex cultural and social means of managing renewable resources and regulating access by their members in a way that ensures resource use does not exceed renewability. Traditions of temporal and spatial rotation of harvesting activities, limited entry regulation, often in the context of communal tenure systems, religious practices, and other customs are ways of reducing exploitation pressures while maintaining the viability of the society as a whole.[45] The

---

ECONOMIC DEVELOPMENT 13-15 (1982) [hereinafter GOODLAND]; TRADITIONAL ECOLOGICAL KNOWLEDGE: A COLLECTION OF ESSAYS (R.E. Johannes ed., 1989); CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 73-74; THE DIVERSITY OF LIFE, *supra* note 9, at 285, 291, 321-22.

43. For a specific society and place, culture is a resource in itself because through culture, environments are conceptually constituted, the means and controls of exploitation are organized, and cumulative resource knowledge stored, taught and used. The cultural resources of indigenous peoples are based on hundreds, often thousands, of years of empirical experience with the ecology of living resources in specific environments.

Nietschmann, *supra* note 42, at 334.

44. *E.g.*, Kenneth I. Taylor, *Deforestation and Indians in Brazilian Amazonia*, *in* BIODIVERSITY, *supra* note 16, at 138, 140; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 83.

45. There is an ever-expanding literature analyzing the successes and failures of various forms of community self-organization in the management of biological resources. These analyses draw a sharp distinction between "open access" resources, which are not protected by effective means of exclusion or limitations on use, and "communal property" resources, which are held in common by a group but collectively managed to prevent over-exploitaiton. While true "open access" conditions can result in a "tragedy of the commons" (Garrett Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968)), communal property institutions of some communities are remarkably successful in conserving biological diversity and maintaining ecosystem stability, while harvesting sufficient resources to ensure survival of the society. With their sophisticated methods for shifting and overlapping uses that adjust to ecological conditions, many of these common property arrangements are quite different from the fixed rights to real property prevalent in industrialized societies, and they are often unrecognized as "property" in the law of the countries

strategies developed to govern human behavior in relation to the ecosystem serve to protect and foster biological diversity in ways that are tailored to the particular environment and support long-term survival of the society.

(d.) *Values.*—Many traditional societies adhere to cultural, religious, or spiritual values that reflect a sense of permanent affiliation with a particular place, responsibility to future generations, and a willingness to forego short-term benefits for long-term viability of the environment. In economic terms, their discount rates are low. Their norms lead them to conserve biological diversity so that resources remain renewable indefinitely for future generations.[46]

All of these characteristics of communities that successfully conserve biological diversity in their immediate environments and live sustainably in one area for many generations make them, from the "environmental" perspective, important actors and allies in the preservation of biological diversity.

How does a recognition of the distinctive contributions of local communities fit within an emerging international environmental law that treats biological diversity as a "global resource"? The status and authority of local communities has become an issue in a broader discussion about how states must exercise sovereign authority in order to meet their trusteeship responsibilities. In essence, the emergence of international environmental norms has brought a scrutiny of states' resource management systems, property rights allocations, and methods of government decisionmaking. If states are seen as trustees of a common heritage, answerable to the international community, then their economic and political systems are subject to international criticism in situations where biological diversity has not been adequately protected.[47] By championing roles for local communities as "participants" in government decisionmaking, as the "stewards" or "custodians" of biological diversity, or as resource "owners" entitled to exclude incompatible uses, advocates of international environmen-

---

where they are found. THE STRUGGLE FOR THE LAND, *supra* note 39; ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION 58-102 (1990) [hereinafter ELINOR OSTROM]; COMMON PROPERTY RESOURCES: ECOLOGY AND COMMUNITY-BASED SUSTAINABLE DEVELOPMENT (Firket Berkes ed., 1989); THE QUESTION OF THE COMMONS: THE CULTURE AND ECOLOGY OF COMMUNAL RESOURCES (B.J. McCay & J.M. Acheson eds., 1987). *See also*, Carol Rose, *The Comedy of the Commons: Custom, Commerce, and Inherently Public Property*, 53 U. CHI. L. REV. 711, 723, 743-44 (1986) (addressing, *inter alia*, the role of custom in managing common resources).

46. ELINOR OSTROM, *supra* note 45, at 35, 88-89. Their values contrast with the discount rates typically applied by government economic planners and corporations. GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 48. *Cf.*, Bryan G. Norton, *The Cultural Approach to Conservation Biology*, *in* CONSERVATION FOR THE TWENTY-FIRST CENTURY 241-46 (David Western & Mary C. Pearl eds., 1989).

47. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 47-52; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 16-18, 37-54.

tal protection for biological diversity have sought, in essence, to influence or to require delegation of government authority in a way that will halt the practices that destroy resources of global importance.[48] This approach is premised on a convergence between local aims and international goals of maintaining ecosystem functions and biological diversity.[49] The insistence on recognition of local rights becomes a method for ensuring that states meet their trusteeship obligations to the international community.

A general recognition of the value of local knowledge, views, and practices in preserving biological diversity may be seen in a growing insistence on opportunities for public participation in government decisionmaking. The World Charter for Nature adopted by the United Nations General Assembly in 1982 called for public input as an integral part of implementing a global conservation strategy:

> All persons, in accordance with their national legislation, shall have the opportunity to participate, individually or with others, in the formulation of decisions of direct concern to their environment, and shall have access to means of redress when their environment has suffered damage or degradation.[50]

Commentators have noted an increasingly wide acceptance of "environmental impact assessment" (EIA) procedures, modeled after the National Environmental Policy Act[51] in the United States, including provisions for public participation in government decision-

---

48. "Returning a measure of control over public lands and resources to local communities is thus fundamental to slowing biodiversity loss in many threatened ecosystems. Such restitution is particularly appropriate in the biologically rich ancestral domains of the world's indigenous peoples." GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 80. *See* Judith Kimerling, *Disregarding Environmental Law: Petroleum Development in Protected Natural Areas and Indigenous Homelands in the Ecuadorian Amazon*, 14 HASTINGS INT'L & COMP. L. REV. 849, 900-03 (1991) [hereinafter Kimerling]; Wirth, *supra* note 7. As an extension of this rationale, environmental groups have made increasingly direct use of human rights arguments, see discussion *infra* part I.B. *See* Kimerling, *supra*, at 889 n. 142 and accompanying text.

49. "Native populations and national resource managers are appropriate allies . . . . Given . . . the close union of the goals of native people to preserve the environment in perpetuity with the goals of the advocates of protected areas, alliance is a logical step." Leslie A. Brownrigg, *Native Cultures and Protected Areas: Management Options* [hereinafter Brownrigg], *in* CULTURE AND CONSERVATION: THE HUMAN DIMENSION IN ENVIRONMENTAL PLANNING 33, 36 (Jeffrey A. McNeely & David Pitt eds., 1985). *See also* James C. Clad, *Conservation and Indigenous Peoples: A Study of Convergent Interests*, *in* CULTURE AND CONSERVATION: THE HUMAN DIMENSION IN ENVIRONMENTAL PLANNING at 45, 46-47 [hereinafter Clad] ("The coincidence of interests characterizing the indigenous peoples' movement and the international lobby for better management of natural resources has been apparent for some time.").

50. World Charter for Nature, *supra* note 29, Principle 23.

51. 42 U.S.C. §§ 4321-4370 (1988).

making.[52] While such procedural requirements do not grant any particular weight to the views of local communities, they do provide an avenue for opposing environmentally destructive practices and seeking protection of existing uses of biological resources.

The logic of granting more extensive protections for local communities as an aspect of protecting biological diversity has been increasingly well-articulated in nature conservation literature, in parks administration policies, and in the policies of international lending institutions.[53] In part, these developments reflect a sharpened focus on the roles of local communities from an "environmental" perspective. At the same time, though, the local community/environment relationship has figured prominently in developments in international human rights law. This "human rights" perspective on local communities and their environments is the subject of the following section.

B. A Human Rights Perspective: Biological Diversity as a Local Resource Integral to Community Identity and Survival

From the "environmental" perspective just examined, biological diversity is a global resource of interest to the international community, and recognition of local communities' roles is a means for protecting that resource. This Article turns now to examples of a rather different, "human rights" perspective on local communities which has evolved simultaneously. Human rights documents have addressed the same problems of resource exploitation and ecological destruction, but from another angle. Rather than focusing on notions of global interdependence and obligations to an international community, the "human rights" perspective has emphasized states' duties

---

52. Nicholas A. Robinson, *The 1991 Bellagio Conference on U.S.-U.S.S.R. Environmental Protection Institutions: International Trends in Environmental Impact Assessment*, 19 B.C. ENVTL. AFF. L. REV. 591 (1992); Remarks by David Wirth *in Environment, Economic Development and Human Rights*, *supra* note 8, at 45-50 (discussing, *inter alia*, the Goals and Principles of Environmental Impact Assessment adopted by the United Nations Environment Program in 1987, U.N. Doc. UNEP/WG.152/4 Annex III (1987), *reprinted in* 17 ENVT'L POL'Y & L. 36 (1987), *adopted* G.C. Dec. 14/25 (1987), U.N. GAOR 42nd Sess., Supp. No. 25, at 77, U.N. Doc. A/42/25 (1987). *See also* Ksentini Preliminary Report, *supra* note 8, paras. 18, 81-84. Public participation through environmental impact assessment procedures has become a central recommendation of organizations involved in nature conservation.*See* CARING FOR THE EARTH, *supra* note 13, at 66-67; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 69; Kimerling, *supra* note 48, at 900-03; Sierra Club Legal Defense Fund, *supra* note 8, at 45-80.

53. BROWNRIGG, *supra* note 49, at 37-43; GOODLAND, *supra* note 42; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 79-96; *Environment, Economic Development and Human Rights*, *supra* note 8; Madhav Gadgil, *Conserving Biodiversity as if People Matter: A Case Study from India*, XXI Ambio 266 (1992); Jeffrey A. McNeely, *supra* note 35; Wirth, *supra* note 7, at 2649 nn. 16-17, 2663 n. 60.

to protect local communities' survival and autonomy. If the "environmental" approach has treated human communities "objectively" as aspects of a natural ecological system in need of international environmental protection, then the "human rights" approach, conversely, may be said to treat the environment "subjectively" as an inseparable aspect of a community's cultural identity that should be protected under international human rights law.

The close affiliation with the natural environment that makes some communities so adept at preserving biological diversity and surviving within the constraints of local renewable resources may also render precarious their situation in the larger political and economic order. Communities that depend for most of their needs on the renewable resources of local ecosystems are vulnerable in the face of environmental degradation. The clearing of a forest or the pollution of a river may mean physical catastrophe for members of a community with a subsistence economy, and such environmental destruction may also doom cultural traits and practices that are tied to, and grow out of, the community's relationship with its habitat.

The examples of communities drastically affected by environmental destruction prominently include many tribes in the countries of the Amazon, the inhabitants of the Chittagong Hill Tracts in Bangladesh, and forest dwellers in Indonesia and Malaysia, to name a few.[54] The precise causes and effects of the environmental destruction vary among different places and communities, but some prevalent scenarios merit attention. Particularly vivid cases of human suffering and conflict involve large-scale resource exploitation for commercial gain. The exploitation may take the form of mining, logging, hydroelectric dams, agricultural operations, or commercial fishing, but the endeavors have similar effects, destroying biological resources and impairing longstanding subsistence uses, whether through pollution, burning or cutting of vegetation, inundation, or over-harvesting.[55]

---

54. Documentation of effects on indigenous peoples has been painstakingly summarized in the course of United Nations-sponsored studies. An extensive inquiry was conducted by Special Rapporteur Jose R. Martinez Cobo for the United Nations Sub-Commission on Prevention of Discrimination and Protection of Minorities and summarized in a series of reports in 1981 through 1983. J. Martinez Cobo, Study of the Problem of Discrimination against Indigenous Populations, *reissued as* U.N. Doc. E/CN.4/Sub.2/1986/7 and Adds.1-4 [hereinafter Indigenous Study]. Such effects are also the subject of ongoing consideration and documentation by the Working Group on Indigenous Populations, of the Sub-Commission on Prevention of Discrimination and Protection of Minorities. *See infra*, notes 75-86 and accompanying text. Studies of specific peoples and areas, with references to other literature, may also be found on a continuing basis in the issues of Cultural Survival Quarterly.

55. *E.g.*, SUSANNA HECHT & ALEXANDER COCKBURN, THE FATE OF THE FOREST (1990) [hereinafter SUSANNA HECHT & ALEXANDER COCKBURN] (on mining and on destruction of forest for agriculture); C. Patrick Morris, *Hydroelectric Development and the Human Rights of Indigenous People*, *in* STRUGGLE FOR THE LAND, *supra*

Often such projects have been sponsored or subsidized by national governments or international funding agencies.[56]

In some instances, the state simply does not recognize the existence of local rights encompassing the uses that have long been made of ecosystems; land ownership is assigned instead to the government or to other private enterprises. In other instances, the law provides protection of local customary uses in principle, but subordinates these uses to other forms of exploitation deemed to be in the national interest. Elsewhere the legal system simply fails to protect a community's rights with adequate enforcement mechanisms. In all of these situations, however, the benefits of exploitation flow to individuals and institutions that do not pay the full costs of their activities. The circumstances are often correctly described as a "tragedy of the commons," in which the failure of restrictive mechanisms results in conditions of open access, over-exploitation of resources, and imposition of costs on persons who do not reap the benefits.[57]

The "environmental" perspective, as we saw earlier, seeks to protect the international community at large from harm to its interests in resources of global importance. The "human rights" perspective considers the same phenomena at a different scale, emphasizing the costs imposed on local groups when the resources they depend upon are demolished. It sees local communities as marginalized, hurt and destroyed through an alteration of their environment, or a forcible separation from it.

While the emergence of a generalized "right to the environment" and "right to development" has been a subject of much recent discussion and debate,[58] here this Article focuses on some specific developments in international human rights law that have drawn a direct and explicit connection between preserving the viability of ecosystems and protecting distinctive cultures. The documents discussed below illustrate a "human rights" perspective on defining rights to biological resources that influenced the formulation of

---

note 39, at 193-209 (on dam building); PUBLIC POLICIES AND THE MISUSE OF FOREST RESOURCES (Robert Repetto and Malcolm Gillis eds., 1988) [hereinafter MISUSE OF FOREST RESOURCES] (on deforestation); Kimerling, *supra* note 48 (on oil drilling).

56. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 47-49; MISUSE OF FOREST RESOURCES, *supra* note 55, *passim*; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 16-18.

57. CONSERVING THE WORLD'S BIOLOGICAL DIVERSITY, *supra* note 9, at 47-49; GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 16-17. For a discussion of "commons" management, see *supra* note 45, distinguishing "open access" commons on the one hand and successfully managed common pool resources on the other. In a typical scenario, open access conditions are accompanied by a breakdown in common property institutions. *See generally*, ELINOR OSTROM, *supra* note 45.

58. *See supra* note 8. *See also* Note, *International Human Rights Law and the Earth: The Protection of Indigenous Peoples and the Environment*, 31 VA. J. INT'L L. 479 (1991).

provisions in the UNCED documents concerning indigenous peoples' communities and other local communities.

1. *The International Labor Organization's Conventions Concerning Indigenous and Tribal Peoples.*—The International Labor Organization (ILO) gave early and notable attention to the vulnerability of specific human communities to disruption of their relationship to the lands they inhabit. The ILO adopted Convention No. 107 (Convention Concerning the Protection and Integration of Indigenous and Other Tribal and Semi-Tribal Populations in Independent Countries[59]) in 1957. It issued a revised Convention No. 169 (Convention Concerning Indigenous and Tribal Peoples in Independent Countries[60]) in 1989. These two Conventions have dealt directly with the rights of communities defined as "indigenous" or "tribal" to remain in and manage the lands that they have traditionally used and occupied.[61]

The provisions of Convention No. 107 addressing rights to land are oriented primarily toward guaranteeing property rights that reflect a community's use and dependence upon natural resources. Thus it provides that the "right of ownership, collective or individual, of the members of the populations concerned over the lands which these populations traditionally occupy shall be recognised."[62] It also protects the self-management of the group with respect to the land: "Procedures for the transmission of rights of ownership and use of

---

59. Convention Concerning the Protection and Integration of Indigenous and Other Tribal and Semi-Tribal Populations in Independent Countries, June 26, 1957, 328 U.N.T.S. 247 [hereinafter Convention No. 107].

60. International Labour Organisation: Convention Concerning Indigenous and Tribal Peoples in Independent Countries, adopted June 27, 1989, *reprinted in* 28 I.L.M. 1382 (1989) [hereinafter Convention No. 169]. Convention No. 169 closed Convention No. 107 to further signatures.

61. Convention No. 107, *supra* note 59, art. 1, 328 U.N.T.S. at 251 ("tribal or semi-tribal populations"); Convention No. 169, *supra* note 60, art. 1 ("indigenous" and "tribal" "peoples"). The task of defining the relevant communities has been a source of considerable controversy in the drafting of Convention No. 169 and other documents. *See* HURST HANNUM, AUTONOMY, SOVEREIGNTY, AND SELF-DETERMINATION: THE ACCOMMODATION OF CONFLICTING RIGHTS 88-91 (1990) [hereinafter HURST HANNUM]; NATAN LERNER, GROUP RIGHTS AND DISCRIMINATION IN INTERNATIONAL LAW 100-03 (1991). The controversy relates to the question of whether the community will be self-defining, or defined by others. Convention No. 169 considers "self-identification" to be a "fundamental criterion." Convention No. 169, *supra* note 60, art. 1, para. 2.

62. Convention No. 107, *supra* note 59, art. 11, 328 U.N.T.S. at 256. Removal of populations from their traditional lands is prohibited except in the interests of "national security," "national economic development," or "health" of the population. *Id.* art. 12, 328 U.N.T.S. at 256. Populations that are removed from their lands should receive comparable lands, "of quality at least equal to that of the lands previously occupied by them, suitable to provide for their present needs and future development." *Id.* art. 12, 328 U.N.T.S. at 256-58. The Convention allows alternative compensation "in money or in kind" if the populations concerned "prefer" it. *Id.* art. 12, 328 U.N.T.S. at 258.

land which are established by the customs of the populations concerned shall be respected."[63] Each of these guarantees is broadly limited by provisions recognizing various overriding state interests, as well as state authority in determining the interests of the populations. The effectiveness of the convention has been criticized, as a consequence;[64] the convention did, however, establish unprecedented international recognition for the land rights of groups within states.[65]

Though not couched in environmental terms, the guarantees set forth in Convention No. 107 are notable for their implicit acknowledgement of an ecological dependence of communities on their environment. Underlying the convention's expressed goal of protecting rights to land is a recognition of the reliance of these groups on the healthy functioning of ecosystems, and their extreme vulnerability in the face of environmental destruction.

These themes were more fully expressed and developed through efforts in the 1980s to revise and update the convention. Convention No. 169, as finalized in 1989, made significant additions and deletions to the language of Convention No. 107.[66] The language of the new convention removes the "assimilationist" orientation of the earlier version,[67] deleting provisions that contemplate integration of peoples into a dominant society and placing a greater emphasis on rights of peoples to be and remain culturally different and to take control over matters important to the group, free of state interference and encroachments by others.[68]

Among the purposes of this autonomy is the protection of a community's distinct relationship to its environment. The revisions in the convention bring an explicit ecological rationale to the defi-

---

63. *Id.* art. 13, 328 U.N.T.S. at 258. This right is qualified by the proviso that these customs must fit "within the framework of national laws and regulations, in so far as they satisfy the needs of these populations and do not hinder their economic and social development." *Id.*

64. *See* Andrée Lawrey, *Contemporary Efforts to Guarantee Indigenous Rights Under International Law*, 23 VAND. J. TRANSNAT'L L. 703, 717-20 (1990).

65. HURST HANNUM, *supra* note 61, at 76-77, 92-93 (1990); Robert A. Williams, *Encounters on the Frontiers of International Human Rights Law: Redefining the Terms of Indigenous Peoples' Survival in the World*, 1990 DUKE L.J. 660, 664-65, 672-76 (1990) [hereinafter Williams].

66. For extensive discussion of the revisions to Convention No. 107, see Howard R. Berman, *The International Labour Organization and Indigenous Peoples: Revision of I.L.O. Convention No. 107 at the 75th Session of the International Labour Conference, 1988*, 41 INT'L COMM'N JURISTS REV. 48 (1988).

67. Convention No. 169, *supra* note 60, Preamble.

68. "The peoples concerned shall have the right to decide their own priorities for the process of development as it affects their lives, beliefs, institutions and spiritual well-being and the lands they occupy or otherwise use . . . ." *Id.* art. 7, para. 1. The term "populations" is also replaced by "peoples," a term suggestive of broader rights to self-determination, although any implication of international rights in the use of this term is expressly denied. *Id.* art. 1, para. 3.

nition of rights, calling for special measures to safeguard "cultures and environment."[69] The convention also requires governments to "respect the special importance for the cultures and spiritual values of the peoples concerned of their relationship with the lands or territories . . . which they occupy or otherwise use, and in particular the collective aspects of this relationship."[70] The new provisions go beyond the recognition of traditional property rights to land, introducing protection from environmental impacts in general,[71] expanding the term "lands" to include "the total environment" of the community[72] and explicitly noting that rights may extend into areas not exclusively occupied by the peoples concerned.[73] Overtly transcending definitions of exclusive property rights drawn by fence lines, the convention mandates a translation of overlapping human uses and dependencies into the formulation of rights, even if existing national law offers no adequate categories.

Convention No. 169 also introduces new procedural safeguards to ensure implementation of the protections mandated. It grants rights to "participate," to obtain "effective representation," and to "be consulted" in government decisions affecting the relationship of communities to their environment.[74] It also requires effective measures by the government to delineate and enforce a community's rights to ownership and possession.[75] Underlying these guarantees of ownership

---

69. *Id.* art. 4. These topics have been added to the list of "persons, institutions, property, [and] labour" originally provided in Convention No. 107, *supra* note 59, art. 3, para. 1, 328 U.N.T.S. at 252.

70. Convention No. 169, *supra* note 60, art. 13, para. 1.

71. *See id.* art. 7, paras. 3, 4 ("Governments shall ensure that, whenever appropriate, studies are carried out, in co-operation with the peoples concerned, to assess the social, spiritual, cultural and environmental impact on them of planned development activities . . . . Governments shall take measures, in co-operation with the peoples concerned, to protect and preserve the environment of the territories they inhabit.").

72. "'[L]ands' . . . shall include the concept of territories, which covers the total environment of the areas which the peoples concerned occupy or otherwise use." *Id.* art. 13, para. 2. The new convention does not guarantee title to subsurface resources, however. Mineral rights may be held by the state, although indigenous peoples should receive compensation for damages from mining activities, and "wherever possible" participate in the benefits of exploitation. *Id.* art. 15, para. 2.

73. "[M]easures shall be taken in appropriate cases to safeguard the right of the peoples concerned to use lands not exclusively occupied by them, but to which they have traditionally had access for their subsistence and traditional activities. Particular attention shall be paid to the situation of nomadic peoples and shifting cultivators in this respect." *Id.* art. 14.

74. *See id.* arts. 15, 16, 17, 23.

75. "Governments shall take steps as necessary to identify the lands which the peoples concerned traditionally occupy, and to guarantee effective protection of their rights of ownership and possession." *Id.* art. 14, para. 2.

and political voice is a recognition that the well-being and cultural survival of the protected communities closely depend on the preservation and use of renewable resources in the local environment, and that the scope of this dependence needs to be translated into property rights and other government guarantees.[76]

While the conventions of the International Labor Organization (ILO) remain the only binding international human rights documents directly addressing these issues, parallel efforts over the last decade of the Working Group on Indigenous Populations of the United Nations Sub-Commission on Prevention of Discrimination and Protection of Minorities have also addressed and extended these themes, in the course of developing a Declaration on the Rights of Indigenous Peoples. As discussed in the following section, the process of formulating a draft declaration has produced not only a reiteration of the ideas expressed in the ILO conventions, but also further consideration of the culture-environment relationship, with an increased focus on expanding access to biological resources, excluding incompatible uses, and fostering the autonomy of a community in making decisions about ecosystem management.

2. *The Draft Universal Declaration on the Rights of Indigenous Peoples of the Working Group on Indigenous Populations.*—The Working Group on Indigenous Populations, an organ of the United Nations Sub-Commission on Prevention of Discrimination and Protection of Minorities,[77] has been working since 1982 to develop a declaration on indigenous rights for submission to the General Assembly. Its work has drawn upon the extensive Study of the Problem of Discrimination against Indigenous Populations[78] prepared previously for the Sub-Commission by Special Rapporteur Jose R. Martinez Cobo, and upon its own investigations. The Working Group's sessions have become a central international human rights forum for indigenous peoples to gain recognition of issues affecting survival of their communities and cultures.[79]

---

76. "[S]ubsistence economy and traditional activities of the peoples concerned, such as hunting, fishing, trapping and gathering, shall be recognised as important factors in the maintenance of their cultures and in their economic self-reliance and development." *Id.* art. 23, para. 1.

77. The five members of the Working Group on Indigenous Populations are appointed from the United Nations Sub-Commission on the Prevention of Discrimination and Protection of Minorities, a subsidiary body of the Commission on Human Rights of the United Nations Economic and Social Council (ECOSOC).

78. Indigenous Study, *supra* note 54. The study is of special interest in the evolution of legal ideas regarding ecological relationships because of its extensive investigation of land rights of indigenous populations in countries around the world and because of its documentation of indigenous populations' efforts to gain legal control of their environments. See in particular the chapter on land, U.N. Doc. E/CN.4/Sub.2/1986/7/Add.3 at 4-205, and the recommendations and conclusions on land rights, *id.* Add.4 at 15-19, 39-42.

79. Russel L. Barsh, *Current Developments: Indigenous Peoples: An Emerg-*

Of relevance to the current discussion is the consideration that the Working Group has given to the rights of indigenous peoples to the land that they have traditionally used and occupied, and to other aspects of their environment. The draft language for a Declaration on the Rights of Indigenous Peoples, resulting from the Working Group's 1992 session, reflects the focus of its ongoing endeavors.[80] The Working Group has given particular attention to the relationship of indigenous peoples to their immediate environments and to the articulation of rights to maintain this relationship. The recent draft provides: "Indigenous peoples have the right to recognition of their distinctive and profound relationship with the total environment of the lands, territories and resources which they have traditionally occupied or otherwise used."[81]

More explicitly than the ILO conventions, the Working Group's draft language gives attention to the ways in which cultural identity as well as physical well-being are tied to environmental preservation. "Indigenous peoples have the collective and individual right to be protected from cultural genocide, including the prevention of and redress for: . . . [d]ispossession of their lands, territories or resources . . . ."[82] Proposed language would protect the cultural aspects of the environmental relationship not merely by recognizing property rights in land and other resources, but also by expanding definitions of intellectual property to include the community's accumulated knowledge about the uses of its environment.[83]

---

ing Object of International Law, 80 AM. J. INT'L L. 369 (1986) and Williams, *supra* note 65, are insightful articles that address the importance of the Working Group proceedings in providing an ongoing international forum for addressing the rights of indigenous peoples and the terms of their existence in the countries they inhabit. *See also* Russel L. Barsh, *Indigenous North America and Contemporary International Law*, 62 OR. L. REV. 73 (1983).

80. Preambular and Operative Paragraphs of the Draft Declaration as agreed upon by the Members of the Working Group at First Reading [hereinafter Draft Declaration], Annex I to Discrimination against Indigenous Peoples: Report of the Working Group on Indigenous Populations on its tenth session, U.N. Doc. E/CN.4/Sub.2/1992/33 (1992) [hereinafter Tenth Session Report].

81. Draft Declaration, *supra* note 80, operative para. 15. This right is elaborated in other paragraphs of the draft declaration: "Indigenous peoples have the collective and individual right to own, control and use the lands and territories they have traditionally occupied or otherwise used." *Id.* operative para. 15. "Indigenous peoples have the right to the protection and, where appropriate, the rehabilitation of the total environment and productive capacity of their lands and territories . . . ." *Id.* operative para. 18. "In no case may indigenous peoples be deprived of their means of subsistence." *Id.* operative para. 21. The draft also proposes a right to restitution or fair compensation for lands taken away. *Id.* operative para. 17.

82. *Id.* operative para. 7.

83. The draft thus would grant indigenous peoples "the right to special measures for protection, as intellectual property, of their traditional cultural manifestations, such as . . . seeds, genetic resources, medicine and knowledge of the useful properties of fauna and flora." *Id.* operative para. 19.

Questions of local autonomy in the management of the community's environment have also been addressed. In these respects, too, the Working Group has gone well beyond the provisions of Convention No. 169 in proposing relinquishment of state control over management decisions and protection of local management systems from competing claims to authority, whether by a government or by private parties asserting some superior property right. The recent draft would establish an explicit "right of self-determination" on the part of indigenous peoples,[84] and provide, in the context of natural resources management, that the state should not interfere in, and should affirmatively protect, the community's own forms of control. Indigenous peoples' right to ownership, control, and use of resources would thus include the right to "the full recognition of their own laws and customs, land-tenure systems and institutions for the management of resources, and the right to effective measures by States to prevent any interference with or encroachment upon these rights."[85]

Proposed provisions would also acknowledge indigenous peoples' "right to autonomy in matters relating to their own internal and local affairs, including . . . land and resources administration, environment and entry by non-members."[86] Implicitly, the community's relationship to its environment is changing rather than static. Environmental alteration caused by human activity is to take place through decisionmaking within the group, however, rather than by outside fiat.

To summarize, the ILO conventions and the Working Group's efforts at drafting a declaration have reflected a "human rights" perspective on the conservation of biological diversity, in the specific context of protecting communities of indigenous people. While this abbreviated discussion does not comprehensively review the emergence of indigenous peoples' rights, it serves to highlight themes of local community empowerment and protection that became influential at UNCED.

As the documents selected above illustrate, the "human rights" approach seeks secure property rights, as well as rights to self-governance and rights to participate in government decisionmaking, as means for affirming communities' particular relationships with the environment. This approach focuses on maintaining a community's access to biological resources, excluding incompatible uses by others,

---

84. *Id.* operative para. 1. By virtue of this right "they may freely determine their political status and institutions and freely pursue their economic, social, and cultural development. An integral part of this is the right to autonomy and self-government." *Id.*

85. *Id.* operative para. 16.

86. *Id.* operative para. 27.

and fostering community autonomy. Though this approach uses, in part, the language of property rights, the "property" at issue is so closely tied to community viability that it is treated as an integral aspect of life and identity, which may not be taken away except in exceptional circumstances with special provisions for restitution.[87] Protection of biological diversity in the environment becomes a necessary part of protecting the existence and distinctive culture of the community.

The focus and emphasis of the "human rights" perspective differ from the "environmental" perspective described earlier. The legal terms that have given expression to the two approaches might indeed appear contradictory at first glance. The "environmental" perspective contends that biological diversity and the health of ecosystems must be managed and protected in the interests of an international community, while the "human rights" perspective asserts the interests of a local community. The language of ownership and authority found in the two approaches is distinct. The "environmental" approach demands unified, international controls to guide resource management, while the "human rights" approach demands property rights, local autonomy, self-determination, and protection from outside encroachment.

Nevertheless, as we have seen, the two perspectives converge on similar conclusions about the need to prevent destructive exploitation of resources by governments or private entities, and the need to use biological resources sustainably. Perhaps because existing international law has seemed insufficiently effective, advocates for "environmental" and "human rights" causes have reached out, in recent years, to form alliances and to make use of each other's legal rhetoric to gain additional leverage. Increasingly, environmental organizations endorse local communities' "rights"[88] while indigenous peoples and other local communities adopt the language of international environmental protection, assuming the role of "trustees" on behalf of a broader community.[89]

---

87. *Cf.* Margaret J. Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849, 1904-05 (1987) ("We tend to view things internal to the person as inalienable and things external as freely alienable . . . . A better view of personhood . . . does not conceive of the self as pure subjectivity standing wholly separate from an environment of pure objectivity . . . . Contextuality means that physical and social contexts are integral to personal individuation, to self-development.").

88. *See* Kimerling, *supra* note 48; Sierra Club Legal Defense Fund, *supra* note 8; Fran Spivy-Weber, *The Missing Forest Principles*, *in* NETWORK '92, Jan. 1992, at 3 (on file with *Tennessee Law Review*) (editorial on behalf of National Audubon Society, advocating stronger protections for rights of indigenous peoples and forest dwellers). *See also* Wirth, *supra* note 7, at 2651 (noting how the stake of a local group in protecting a particular habitat can lend legitimacy to endeavors of U.S. environmental organizations).

89. "Our accumulated knowledge about the ecology of our home, our models

C. *A Synthesis of "Human Rights" and "Environmental" Perspectives: The World Commission on Environment and Development's View of the Role of Local Communities in "Sustainable Development"*

The World Commission on Environment and Development, which was established by the United Nations General Assembly in 1984 to study problems of environmental protection and economic development and to recommend new forms of international cooperation, provided an important forum that gave momentum to a synthesis of the "environmental" and the "human rights" perspectives. The Commission's report, *Our Common Future*,[90] firmly joined the goals of environmental protection and economic development in its recommendations, establishing "sustainable development" as the dominant theme for subsequent international environmental policy-making.[91] Echoing ideas expressed earlier in the 1980 *World Conservation Strategy* published by the International Union for Conservation of Nature and Natural Resources, the United Nations Environment Programme, and the World Wildlife Fund,[92] the report defined sustainable development to mean "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."[93]

---

for living within the Amazon Biosphere, our reverence and respect for the tropical forest and its other inhabitants, both plant and animal, are the keys to guaranteeing the future of the Amazon Basin, not only for our peoples, but also for all of humanity." Coordinating Body for the Indigenous Peoples' Organizations of the Amazon Basin (COICA), To the Community of Concerned Environmentalists: Our Agenda 1 (1989) (on file with *Tennessee Law Review*). "We rubber tappers demand to be recognized as producers of rubber and as the true defenders of the forest." Platform of the National Rubber Tappers' Council, October 11-17, 1985, *reprinted in* SUSANNA HECHT & ALEXANDER COCKBURN, *supra* note 55, Appendix E, at 262. *See also* Discrimination against Indigenous Peoples: Report of the Working Group on Indigenous Populations on its ninth session, U.N. Doc. E/CN.4/Sub.2/1991/40/Rev.1 at 10 (1991) (comments presented to the Working Group noting "the traditional role of indigenous peoples as custodians of the environment").

90. OUR COMMON FUTURE, *supra* note 32 (often called the "Brundtland Report" after the commission's chairman Norwegian Prime Minister Gro Harlem Brundtland).

91. See LYNTON KEITH CALDWELL, INTERNATIONAL ENVIRONMENTAL POLICY 207 (1990) for commentary on the repercussions of OUR COMMON FUTURE.

92. INTERNATIONAL UNION FOR CONSERVATION OF NATURE AND NATURAL RESOURCES ET AL, WORLD CONSERVATION STRATEGY: LIVING RESOURCE CONSERVATION FOR SUSTAINABLE DEVELOPMENT (1980) [hereinafter WORLD CONSERVATION STRATEGY].

93. OUR COMMON FUTURE, *supra* note 32, at 8. The definition has been criticized as subject to a range of contradictory interpretations. The more recent update of the *World Conservation Strategy*, CARING FOR THE EARTH, *supra* note 13, at 10, defines sustainable development to mean "improving the quality of human life while living within the carrying capacity of supporting ecosystems." It further defines "sustainable use" to mean "using [renewable resources] at rates within their capacity for renewal." *Id.*

Conjoining international nature preservation goals and human rights goals within the scope of "sustainable development," *Our Common Future* presented a dual perspective on the role of local communities. On the one hand, the report emphasized the value to the international community of local communities' extensive knowledge about the ecosystems where they live: "These communities are the repositories of vast accumulations of traditional knowledge and experience."[94] The report found that local communities are particularly qualified to manage biological resources so as to serve global environmental goals: "[Their] traditional life-styles . . . can offer modern societies many lessons in the management of resources in complex forest, mountain, and dryland ecosystems."[95] Cultural destruction caused by environmental degradation is "a loss for the larger society, which could learn a great deal from their traditional skills in sustainably managing very complex ecological systems."[96]

On the other hand, the report simultaneously emphasized the vulnerability of indigenous and tribal communities to environmental degradation or dispossession, urging recognition and protection for communities' rights to land and other resources that sustain their way of life. The report used terms similar to those of the ILO conventions and the draft language of the Working Group on Indigenous Populations.[97]

The report suggested that changes in state development policies and laws governing property rights are essential to both cultural and environmental survival. Recognition of land ownership rights within the terms of existing definitions may be insufficient: "The starting point . . . is the recognition and protection of [communities'] traditional rights to land and the other resources that sustain their way of life—rights they may define in terms that do not fit into standard legal systems."[98] Legal reforms should give local communities "a decisive voice" in decisions about resource management.[99] Adequate recognition of rights must also extend to governmental support for the group's own means of self-regulation: "[T]he recognition of

---

94. OUR COMMON FUTURE, *supra* note 32, at 114.

95. *Id.* at 12.

96. *Id.* at 114-15.

97. The "cultural genocide" idea of the Working Group's draft is clearly expressed in the Brundtland report. The report gives the following commentary on how separation of people from their natural surroundings, whether through dispossession or environmental destruction, may have catastrophic human effects: "Many live in areas rich in valuable natural resources that planners and 'developers' want to exploit, and this exploitation disrupts the local environment so as to endanger traditional ways of life . . . . .Many groups become dispossessed and marginalized, and their traditional practices disappear. They become the victims of what could be described as cultural extinction." *Id.* at 114.

98. *Id.* at 115.

99. *Id.* at 116.

traditional rights must go hand in hand with measures to protect the local institutions that enforce responsibility in resource use."[100]

The work of the World Commission on Environment and Development gave rise to the subsequent General Assembly resolution calling for the U.N. Conference on Environment and Development (UNCED) in 1992.[101] By treating environmental and human rights concepts as unified, the Commission laid important groundwork for a further synthesis in international law. By no means did all of the policy recommendations of the Commission's report gain sufficient consensus among states to emerge as provisions in the documents prepared for UNCED. But the need to internationally recognize and protect a role for indigenous peoples and other local communities in the management of ecosystems, for the sake of both global environmental interests and local human concerns, gained a firm hold in the UNCED negotiations.

II. THE CONVERGENCE OF GLOBAL ECOLOGICAL GOALS AND LOCAL COMMUNITY ROLES AT UNCED

A. The Recognition of Global Concerns and the International Goal of Sustainability in the UNCED Documents

1. *The Reaffirmation of Sovereignty.*—A dynamic interplay among global, national, and local notions of "community" and "ownership" characterize the provisions that emerged in the UNCED documents.

The UNCED documents strongly reassert the premise that states have sovereign authority over their natural resources, in words that might seem to preclude local or global empowerment in the management of biological resources. The Rio Declaration in Principle 2 reiterates and expands on the declaration of sovereign authority over natural resources issued twenty years before at the United Nations Conference on the Human Environment:

> States have, in accordance with the Charter of the United Nations and the principles of international law, the sovereign right to exploit their own resources pursuant to their own environmental and developmental policies, and the responsibility to ensure that activities within their jurisdiction or control do not cause damage to the environment of other States or of areas beyond the limits of national jurisdiction.[102]

---

100. *Id.* at 115-16.

101. G.A. Res. 228, U.N. GAOR, 44th Sess., Supp. No. 49, at 300, Dec. 22, 1989.

102. Rio Declaration, *supra* note 1, Principle 2. Similar language (omitting only the reference to "developmental" policies) appears in Principle 21 of the Stockholm Declaration, *supra* note 17.

Agenda 21, implementing the Rio Declaration with detailed plans for action, confirms again that "States have the sovereign right to exploit their own biological resources pursuant to their environmental policies."[103] The Convention on Biological Diversity[104] and the Forest Principles[105] quote Principle 21 of the Stockholm Declaration verbatim; the Convention on Climate Change contains similar provisions.[106] The documents leave no doubt that state sovereignty remains an important part of the international system for managing natural resources.

Read in isolation, these provisions conjure up an image of autonomous states, acting within well-defined boundaries, limited only by their duty not to cause direct damage outside their own territories. Yet, accompanying the apparently absolute pronouncements of these paragraphs are extensive provisions placing sovereign states within a new web of relations and responsibilities to other entities, even with respect to the management of natural resources within sovereign territory. The language of the various documents addressing global and local concerns frames the assertions of sovereignty, conveying a far more complex vision of multiple, interactive scales and forms of governance.

2. *The Concept of a Global Ecosystem.*—Juxtaposed with the assertions of state sovereignty in the UNCED documents are many provisions expressing quite different notions of boundary delineation for the management of living resources. Global ecological interdependence and the concerns of a unified international community are overarching themes. The Rio Declaration recognizes "the integral and interdependent nature of the Earth, our home,"[107] and calls on states to "conserve, protect and restore the health and integrity of the Earth's ecosystem."[108] Agenda 21 speaks of "interdependence,"[109] of "the ecosystems on which we depend for our well-being,"[110] and of "the life-supporting capacities of our planet."[111] Mountain eco-

103. Agenda 21, *supra* note 2, para. 15.3.

104. Convention on Biological Diversity, *supra* note 5, art. 3. The Convention overrides previous doubts about the sovereign authority of states over genetic resources. *See* GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 43-45, 64 (addressing the "common heritage of humankind" concept of the International Undertaking on Plant Genetic Resources, adopted by the United Nations Food and Agriculture Organization in 1983, Resolution 8/83, U.N. FAO, 22d Sess., Annex, at 50, U.N. Doc. C/83/REP (1983)).

105. Forest Principles, *supra* note 3, Principles/Elements, paras. 1(a), 2(a).

106. Convention on Climate Change, *supra* note 5, Preamble.

107. Rio Declaration, *supra* note 1, Preamble. The Preamble also refers to the need for international agreements that "protect the integrity of the global environmental and developmental system." *Id.*

108. *Id.* Principle 7.

109. Agenda 21, *supra* note 2, para. 2.1.

110. *Id.* para. 1.1.

111. *Id.* para. 5.3.

systems, for example, represent "the complex and interrelated ecology of our planet" and are "essential to the survival of the global ecosystem."[112] Oceans and seas and adjacent coastal areas, including those parts within national jurisdiction, are a single marine environment forming "an integrated whole that is an essential component of the global life-support system."[113] Freshwater resources "are an essential component of the earth's hydrosphere and an indispensable part of all terrestrial ecosystems."[114] Biological diversity needs protection because "[o]ur planet's essential goods and services depend on the variety and variability of genes, species, populations and ecosystems."[115] As a basis for development of international strategies, Agenda 21 calls for scientific investigation to assess "the carrying capacity of the planet Earth" and to gain "a better understanding of land, oceans, atmosphere and their interlocking water, nutrient and biogeochemical cycles and energy flows which all form part of the Earth system."[116]

Similar global concerns are expressed in the other UNCED documents. The Forest Principles note the vital ecological functions of forests[117] and their value "to the environment as a whole."[118] The Convention on Biological Diversity affirms "that the conservation of biological diversity is a common concern of humankind,"[119] while the Convention on Climate Change refers to "the global nature of climate change" and to "the role and importance in terrestrial and marine ecosystems of sinks and reservoirs of greenhouse gases."[120]

In sum, the UNCED documents recognize global ecological interdependence, even with respect to forests, mountains, wetlands, and other coastal and land-based resources located within sovereign territory. These ecosystems are all seen as part of global and regional natural systems that transcend state boundaries. Given this global ecological interdependence, the UNCED documents make important assertions about the appropriate form and scale of human organization. The references to "our planet" and "our home," with their unitary and possessory connotations, become shorthand expressions of international community and authority. These ideas are more fully

---

112. *Id.* para. 13.1.
113. *Id.* para. 17.1.
114. *Id.* para. 18.1.
115. *Id.* para. 15.2.
116. *Id.* para. 35.2.
117. Forest Principles, *supra* note 3, Principles/Elements, para. 4 (The "role [of forests] in protecting fragile ecosystems, watersheds and freshwater resources and as rich storehouses of biodiversity and biological resources and sources of genetic material for biotechnology products, as well as photosynthesis, should be recognized.")
118. *Id.* Preamble, para. (f).
119. Convention on Biological Diversity, *supra* note 4, Preamble.
120. Convention on Climate Change, *supra* note 5, Preamble.

developed through the articulation of an international environmental goal of "sustainability" and the establishment of international mechanisms to oversee implementation.

3. *The International Goal of Sustainability.*—The terms "sustainable development," "sustainable use," "sustainable management" and similar phrases appear again and again in the UNCED documents to describe the responsibilities that governments should undertake to fulfill. Agenda 21, implementing the Rio Declaration, proclaims a "global partnership for sustainable development,"[121] and calls on all governments to adopt national strategies that will "[p]rotect the resource base and the environment for the benefit of future generations" while allowing current economic development.[122] Chapters on land resources,[123] deforestation,[124] lands subject to desertification,[125] mountains,[126] agricultural lands,[127] biological diversity,[128] marine ecosystems,[129] and freshwater resources[130] all articulate goals of sustain-

---

121. Agenda 21, *supra* note 2, paras. 1.1, 1.6. The Rio Declaration refers explicitly to the international objective of sustainable development and to means for achieving it. Rio Declaration, *supra* note 1, Principles 1, 4, 5, 7, 8, 9, 12, 20, 21, 22, 27.

122. Agenda 21, *supra* note 2, para. 8.7.

123. Agenda 21 sets a broad objective of "sustainable and integrated management of land resources", *id.* para. 10.5, calls for formulation of "sustainable management" policies by 1996, *id.* para. 10.5(a), and sets out numerous specific measures and activities to support sustainable land use and management, *id.* paras. 10.6-10.18.

124. *Id.* paras. 11.4(g), 11.13(c), 11.14(a) (setting forth measures to ensure the "sustainable management" of forests).

125. *Id.* paras. 12.17(a), 12.26 (measures to ensure "sustainable management" and "conservation of biodiversity" on lands subject to desertification).

126. *Id.* paras. 13.2 (noting the role of mountains as "storehouses of biological diversity"), 13.15(a) (calling for, *inter alia*, planning and management "to prevent soil erosion, increase biomass production and maintain the ecological balance" by the year 2000).

127. *Id.* para. 14.4 (establishing a program for achieving "sustainable development," "sustainable agriculture," and "sustainable utilization" of genetic resources).

128. *Id.* para. 15.1 (establishing objectives and specific activities "to improve the conservation of biological diversity and the sustainable use of biological resources"), para. 15.3 (confirming that states have "the responsibility to conserve their biodiversity and use their biological resources sustainably"). *See also id.* paras. 16.22-16.23 (biotechnology as a tool for achieving sustainable development, through rehabilitation of degraded ecosystems and landscapes).

129. *Id.* paras. 17.1, 17.70-17.87 (establishing a program for ensuring "sustainable development" and "sustainable use and conservation" of the marine environment).

130. *Id.* para. 18.16 ("Water resources development and management should be planned in an integrated manner, taking into account long-term planning needs as well as those with narrower horizons, that is to say, they should incorporate environmental, economic and social considerations based on the principle of sustainability . . . .").

ability for particular ecosystems. The Forest Principles, similarly, state that forests should be "sustainably managed to meet the social, economic, ecological, cultural and spiritual human needs of present and future generations."[131]

Under the Convention on Biological Diversity, "[s]tates are responsible for conserving their biological diversity and for using their biological resources in a sustainable manner."[132] They must formulate and implement strategies and programs to ensure "conservation and sustainable use" of biological resources;[133] "sustainable use" is defined to mean use "in a way and at a rate that does not lead to the long-term decline of biological diversity, thereby maintaining its potential to meet the needs and aspirations of present and future generations."[134] The Convention on Climate Change also requires the parties to that agreement to "[p]romote sustainable management" and to conserve and enhance the ecosystems—terrestrial, coastal, and marine—that serve as sinks and reservoirs for greenhouse gases.[135]

4. *International Mechanisms for Implementation.*—The Convention on Biological Diversity and the Convention on Climate Change both establish a Conference of the Parties, a Secretariat, and a Subsidiary Body to provide scientific and technological advice, and make other institutional arrangements for implementation, including reporting requirements, dispute resolution mechanisms, and procedures for adoption of protocols.[136] Agenda 21, while it is not a binding instrument, also calls for international institutional arrangements "to ensure and review the implementation of Agenda 21 so as to achieve sustainable development in all countries."[137] These arrangements include creation of a Commission on Sustainable Development empowered to receive national reports, and to review the progress that states make in meeting their commitments under Agenda 21.[138] In short, the various UNCED documents all establish environmental objectives based on a central concept of sustainability, and they set forth international institutional mechanisms directed at ensuring implementation.

---

131. Forest Principles, *supra* note 3, Principles/Elements, para. 2(b).

132. Convention on Biological Diversity, *supra* note 4, Preamble.

133. *Id.* art. 6, para. (a); art. 8, paras. (c), (i).

134. *Id.* art. 2.

135. Convention on Climate Change, *supra* note 5, art. 4, para. 1(d).

136. Convention on Biological Diversity, *supra* note 4, arts. 23-28; Convention on Climate Change, *supra* note 5, arts. 7-14.

137. Agenda 21, *supra* note 2, para. 38.8(a).

138. *Id.* paras. 38.11-38.13. The Commission is to report to the Economic and Social Council. *Id.* para. 38.10. Since Agenda 21 seeks to implement the Rio Declaration and the Forest Principles, *id.* paras. 1.6 and 11.13(e), these documents are likewise encompassed in the Commission's proposed mission.

The various UNCED documents go beyond earlier developments in environmental law[139] by clearly establishing the international stake in the viability of all ecosystems—forests, deserts, coastal waters, mountains, rivers, and lands of all types—and in the biological diversity found within them, and by articulating a standard of sustainability. State sovereignty over natural resources is proclaimed to remain intact, but that sovereignty is placed in the context of new obligations and commitments.[140] While the meaning of "sustainable" remains to be further elaborated through additional development of international law[141] the term clearly suggests the emergence of new responsibilities for states in the management of renewable natural resources. The UNCED documents contemplate state responsibilities that extend beyond controlling activities that have a direct and readily provable impact outside of sovereign territory. The responsibilities might well be characterized as the duties of trusteeship, since resources that a state formally holds need to be managed in a way that benefits not merely its own citizens, but others as well, both in separate countries and in future generations.[142]

B. The Recognition of Local Community Authority in the Management of Biological Resources

The increased centralization of environmental law reflected in the formulation of a standard of sustainability has brought with it, as an integral aspect, an increased recognition of local community authority in the management of biological resources. Under the UNCED documents, international environmental protection becomes a route for vindication of local interests and, in particular, of rights of indigenous peoples and other marginalized groups living within the world's areas of richest biological diversity.

The change in perspective that the UNCED documents represent is highlighted by a comparison of the Rio Declaration and Agenda

---

139. *See supra* notes 17-34, and accompanying text.

140. Nowhere is this shift more apparent than in the Agenda 21 provisions on conservation of biological diversity. There, a new phrase regarding state responsibilities is inserted into a sentence that otherwise tracks the sovereign rights provision of the Stockholm Declaration:

> States have the sovereign right to exploit their own biological resources pursuant to their environmental policies, *as well as the responsibility to conserve their biodiversity and use their biological resources sustainably*, and to ensure that activities within their jurisdiction or control do not cause damage to the biological diversity of other States or of areas beyond the limits of national jurisdiction.

Agenda 21, *supra* note 2, para. 15.3 (emphasis added).

141. Agenda 21 calls for further development of international law on sustainable development, with particular attention to "the delicate balance between environmental and developmental concerns." *Id.* para. 39.1.

142. *See supra* notes 31-34 and accompanying text.

21 to their earlier counterparts, the Stockholm Declaration and its accompanying action plan.[143] The Stockholm documents barely mention a role for groups or communities. Rather, they express a clear vision of governments and international bodies, acting in a modern, technocratic fashion, with ample advice and guidance from scientific experts employed by national and international administrative agencies. Solutions to environmental problems are to be found through careful, centralized planning and management, based on data-gathering and rational analysis. While local entities are acknowledged tangentially as sources of information, they do not figure as central actors in the scheme of natural resources management presented at Stockholm.

Twenty years later, a much-altered international mandate has emerged. We see now, in addition to references to centralized management and technical expertise, an international plan of action that gives a prominent place to individuals, to their communities and organizations, and to relationships among them that bypass the formal structures of government.

1. *Participation and Involvement.*—The new international mandate takes the form, first, of a generalized right to participation in decision-making, backed by public access to government information. The provisions of the Rio Declaration track the basic concepts of citizen participation in environmental impact assessment proceedings in the United States under the National Environmental Policy Act:[144]

> Environmental issues are best handled with the participation of all concerned citizens . . . . At the national level, each individual shall have appropriate access to information concerning the environment . . . including information on . . . activities in their communities, and the opportunity to participate in decision-making processes . . . . Effective access to judicial and administrative proceedings, including redress and remedy, shall be provided.[145]

While this section of the Rio Declaration is focused on individuals, Agenda 21 calls on governments to extend the opportunity for participation not only to individuals but to groups and organizations. Agenda 21 expresses a far-reaching endorsement of participation in decision-making by "all social groups."[146]

---

143. Stockholm Report, *supra* note 17.

144. 42 U.S.C. §§ 4321-4370 (1988). The ideas expressed in the Rio Declaration expand on a similar provision adopted by the General Assembly ten years before, in the World Charter for Nature, *supra* note 29, para. 23. Regarding the emergence of rights to participation, see *supra* notes 8, 50-52 and accompanying text.

145. Rio Declaration, *supra* note 1, Principle 10. *See also id.* Principle 20 (calling specifically for participation of women).

146. Agenda 21, *supra* note 2, specifically identifies as "major groups" women, *id.* ch. 24, children and youth, *id.* ch. 25, indigenous people and their communities,

Agenda 21 links the opportunity for broad public participation by both individuals and groups directly to attainment of the international goal of sustainable development:

> One of the fundamental prerequisites for the achievement of sustainable development is broad public participation in decision-making . . . . This includes the need of individuals, groups and organizations to participate in environmental impact assessment procedures and to know about and participate in decisions, particularly those which potentially affect the communities in which they live and work.[147]

Similarly inclusive statements are found elsewhere throughout Agenda 21[148] and in the Forest Principles.[149]

The generalized right of participation envisioned in the Rio Declaration, Agenda 21, and the Forest Principles is essentially procedural. While these provisions serve to open government decisions to scrutiny by diverse individuals and groups, they do not in themselves give priority to the views of any particular individual or group, and decision-making authority is left in the hands of national governments. Yet the clear expectation of these provisions is that government decision-making will be altered through the consideration of information and views that might otherwise be overlooked and, furthermore, that the procedural avenues established will provide a

---

id. ch. 26, non-governmental organizations, *id.* ch. 27, local authorities, *id.* ch. 28, workers and trade unions, *id.* ch. 29, business and industry, *id.* ch. 30, the scientific and technological community, *id.* ch. 31, and farmers, *id.* ch. 32. These categories overlap: women and indigenous people may also be farmers, for instance, and all of these groups may form non-governmental organizations.

147. *Id.* para. 23.2.

148. *E.g.*, *id.* para. 1.3 ("[t]he broadest public participation and the active involvement of the non-governmental organizations and other groups"), para. 2.6 ("progress towards democratic government . . . allow[ing] for full participation of all parties concerned"), para. 8.3(c) ("mechanisms to facilitate the involvement of concerned individuals, groups and organizations in decision-making"), para. 10.5(d) ("mechanisms to facilitate the active involvement and participation of all concerned, particularly communities and people at the local level, in decision-making on land use and management"), para. 11.2 ("participation of the general public" and various groups, in combating deforestation), para. 12.55 ("go beyond the theoretical ideal of popular participation and . . . focus on obtaining actual active popular involvement" in controlling desertification), para. 13.16 ("[e]nhance popular participation in the management of local resources through appropriate legislation" in managing mountain watersheds), para. 17.129 ("undertake appropriate institutional reforms . . . including . . . community participation in the planning process" for management of marine and coastal areas), para. 18.9(c) ("an approach of full public participation, including that of women, youth, indigenous people, local communities, in water management policy-making and decision-making").

149. "Governments should promote and provide opportunities for the participation of interested parties, including local communities and indigenous people, industries, labour, non-governmental organizations and individuals, forest dwellers and women, in the development, implementation and planning of national forest policies." Forest Principles, *supra* note 3, Principles/Elements para. 2(d).

forum for holding governments to a substantive standard of sustainability. National proceedings therefore are expected to become vehicles for inducing governments to adhere to international goals. Participants in the process who oppose destruction of biological resources are, in some sense, the allies and agents of global environmental interests.

2. *Decentralization.*—A further elaboration of the "grassroots empowerment" theme is found in various provisions that urge decentralization of decision-making. In general, Agenda 21 calls for "[d]elegating planning and management responsibilities to the lowest level of public authority consistent with effective action."[150]

One reason for granting additional authority to local entities, beyond mere participation in national proceedings, is that conservation programs to ensure sustainability of resources over the long term are unlikely to succeed unless the basic needs of people who depend upon the resources for their livelihoods are satisfied.[151] Given the necessity of satisfying local needs and gaining local support in order to attain international environmental protection goals, Agenda 21 calls for a "community-driven" approach to achieving sustainability through programs that integrate local needs with conservation measures.[152]

This approach includes developing programs that will bring about "empowerment of local and community groups through the principle of delegating authority, accountability and resources to the most appropriate level to ensure that the programme will be geographically and ecologically specific."[153] In essence, this proposal for decentralization of government programs suggests that measures for ensuring sustainability may be most effectively designed and carried out locally, with national oversight, rather than centrally, with local input. It does not exempt local action from the central standard of sustainability, but it does grant additional control and independence to local communities in designing a program.

3. *Special Roles for Local Communities.*—Within this broader context of pluralism, decentralization, and "grassroots empowerment," the UNCED documents also single out certain groups for additional attention. The provisions giving special weight to the interests and views of indigenous people, their communities and other local communities are those of relevance to the current discussion.[154]

---

150. Agenda 21, *supra* note 2, para. 8.5(g). *See also*, *id.* paras. 11.4(a), 12.28(a), 14.17(c), 14.18(d), 14.24, 18.12(o).

151. *Id.* para. 3.2.

152. *Id.* paras. 3.7, 3.12.

153. *Id.* para. 3.5(a). *See also id.* para. 32.5 ("The decentralization of decision-making towards local and community organizations is the key in changing people's behavior and implementing sustainable farming strategies.").

154. The Rio Declaration also views "women" and "youth" as groups who

These provisions express the view that the survival of indigenous people in particular, and of some other local communities in addition, is so inextricably tied to the viability of their immediate environments that ensuring their well-being on a local basis necessarily goes hand-in-hand with ensuring the attainment of sustainable development on a global scale.

The Rio Declaration states that

> indigenous people and their communities, and other local communities, have a vital role in environmental management and development because of their knowledge and traditional practices. States should recognize and duly support their identity, culture and interests and enable their effective participation in the achievement of sustainable development.[155]

Thus, the Rio Declaration emphasizes "environmental" reasons for drawing a connection between the conservation of biological resources and the protection of local communities: Certain groups are especially knowledgeable about their environments, adept at conducting sustainable harvesting practices, and in other ways particularly qualified to ensure that biological resources are sustainably managed. Provisions in Agenda 21 emphasize these same points,[156] but also advance a "human rights" perspective: Some local communities are especially vulnerable to destruction of their immediate surroundings because they depend directly upon the living resources of the lands they inhabit in a physical, economic, and cultural way.[157] They have, consequently, both a strong interest in seeing that these resources are sustainably used and conserved, and a special vulnerability to uncontrolled encroachments by others.

The UNCED documents seek to empower these groups in several ways that go beyond opportunities to participate in government decision-making and decentralization in administrative practices. The

---

have particularly important roles to play in achieving sustainable development. Rio Declaration, *supra* note 1, Principles 21, 22. A consideration of the specific provisions relating to these groups is beyond the scope of this Article, but it might be noted that the UNCED documents treat these two groups in somewhat the same manner as it treats local communities, often including women, and sometimes youth, in the same or parallel provisions. Agenda 21 additionally lists various other "major groups" that have special contributions to make. *See supra* note 146. The provisions dealing with these groups tend to highlight their knowledge, existing power or authority, and stake or interest in ensuring implementation of sustainable practices as reasons for encouraging their participation in decision-making.

155. Rio Declaration, *supra* note 1, Principle 23.

156. *E.g.*, Agenda 21, *supra* note 2, para. 26.1 ("Indigenous people and their communities have an historical relationship with their lands . . . . They have developed over many generations a holistic traditional scientific knowledge of their lands, natural resources and environment.").

157. *E.g.*, *id.* para. 26.3(a)(iv). *See also* Convention on Biological Diversity, *supra* note 4, Preamble.

measures delineated fall broadly into three categories. First, a recurrent theme is that the knowledge, traditional practices, values, and cultures of these communities should be given special attention and weight in government proceedings and management plans. In essence, the information that these groups have should not simply be "considered" through ordinary participatory procedures, but affirmatively sought out, recorded, supported, promoted, and incorporated into government management programs.[158] Traditional knowledge and practices, if proven to be effective, should be used as a basis and a model for meeting the substantive standard of sustainability.[159] Even more broadly, governments should protect the values and the cultures of communities that have fostered the accumulation of traditional ecological knowledge and the development of sustainable harvesting practices.[160]

Second, according to the UNCED documents, these local communities should be assured at least a share of the "benefits" derived from the sustainable use of resources in their environments. This concept appears in various forms. The Convention on Biological Diversity states that governments should protect and encourage sustainable traditional uses of biological resources.[161] The Forest Principles conclude that local communities should have an "economic stake" in forest use, and "adequate levels of livelihood and well-being."[162] Agenda 21 generally urges states to provide indigenous people and local communities "wider access to land, water and forest resources."[163] It also calls on coastal states to recognize the "rights"

---

158. "Appropriate indigenous capacity and local knowledge regarding the conservation and sustainable development of forests should . . . be recognized, respected, recorded, developed and, as appropriate, introduced in the implementation of programmes." Forest Principles, *supra* note 3, Principles/Elements, para. 12(d). *See also* Agenda 21, *supra* note 2, para. 26.5 (governments should incorporate indigenous people's "values, views and knowledge" in resource management policies and programs), para. 17.82(c) (governments should promote incorporation of tradition knowledge into fisheries management systems).

159. Thus the Convention on Biological Diversity states that each contracting party shall "[s]ubject to its national legislation, respect, preserve and maintain knowledge, innovations and practices of indigenous and local communities embodying traditional lifestyles relevant for the conservation and sustainable use of biological diversity and promote their wider application . . . ." Convention on Biological Diversity, *supra* note 4, art. 8(j).

160. "National forest policies should recognize and duly support the identity, culture and the rights of indigenous people, their communities and other communities and forest dwellers." Forest Principles, *supra* note 3, Principles/Elements, para. 5(a). *See also* Agenda 21, *supra* note 2, para. 26.6(a).

161. Convention on Biological Diversity, *supra* note 4, art. 10 ("Each Contracting Party shall, as far as possible and as appropriate . . . [p]rotect and encourage customary use of biological resources in accordance with traditional cultural practices that are compatible with conservation or sustainable use requirements . . . .").

162. Forest Principles, *supra* note 3, Principles/Elements, para. 5(a).

163. Agenda 21, *supra* note 2, para. 14.18(b).

of indigenous people and local communities in coastal and marine environments "to utilization and protection of their habitats on a sustainable basis."[164] Implicitly, the protection of traditional uses and the guarantee of continued access to resources mean that incompatible access and exploitation by others should be effectively excluded.

The UNCED documents also speak of allowing local communities to derive economic benefits from uses of traditional knowledge and practices. The Convention on Biological Diversity requires governments to "encourage the equitable sharing of the benefits arising from the utilization of [indigenous and local communities'] knowledge, innovations and practices."[165] Agenda 21, likewise calls for "fair and equitable sharing" of benefits.[166] These provisions do not explain how, and in what amount, the benefits are to be distributed. Nevertheless, the goal, in the context of achieving sustainability, is apparent: one way to enhance sustainable resource management is to reward the knowledge, innovations, and practices that make sustainable use possible. Whether that goal is achieved, for example, by government payments or by new laws delineating marketable intellectual property rights remains to be determined.[167]

Finally, Agenda 21 and the Forest Principles propose that governments consider granting further autonomy and control to communities over resources. Agenda 21 concedes that indigenous people

---

164. *Id.* para. 17.82(b). Agenda 21 also urges that a "right to subsistence" be taken into account in the negotiation and implementation of international fishing agreements. *Id.* para. 17.83. Other provisions speak generically of "measures to ensure that the local population benefits in adequate measure from resource use." *Id.* para. 3.8(c).

165. Convention on Biological Diversity, *supra* note 5, art. 8(j).

166. Agenda 21, *supra* note 2, para. 15.5(e). Agenda 21 similarly calls for "participation" of indigenous people and their communities "in the economic and commercial benefits derived from the use of . . . traditional methods and knowledge." *Id.* para. 15.4(g). It also states, more broadly, that governments should take "incentive measures to encourage the conservation of biological diversity and the sustainable use of biological resources, including the promotion of . . . traditional methods . . . which use, maintain or increase biodiversity." *Id.* para. 15.5(d).

167. Agenda 21 suggests adoption of "appropriate policies and/or legal instruments that will protect indigenous intellectual and cultural property" as optional measures that governments "could" take in order to enhance the ability of indigenous people and their communities to play their role in the attainment of sustainable development. *Id.* para. 26.4(b). Detailed examination of the status of intellectual and cultural property rights for indigenous peoples' knowledge and innovations is beyond the scope of this Article. For a discussion of the issues and relevant legal provisions, see *Recent Developments, Recent Intellectual Property Trends in Developing Countries*, 33 HARV. INT'L L. J. 277 (1992); Discrimination Against Indigenous Peoples: Intellectual property of indigenous peoples: Concise report of the Secretary-General, U.N. Doc. E/CN.4/Sub.2/1992/30 (1992); Diane S. Pickersgill, Indigenous Knowledge and Rainforest Medicinal Plants: Intellectual Property and Natural Resource (April 13, 1992) (unpublished manuscript, on file with *Tennessee Law Review*).

and their communities "may require . . . greater control over their lands, [and] self-management of their resources."[168] To this end, governments "could" consider ratifying ILO Convention No. 169, supporting adoption by the General Assembly of a declaration on indigenous rights, and establishing under national law or policy the right of indigenous people to preserve their own administrative systems.

All of the UNCED documents avoid stating that local communities should necessarily gain full title to the areas that they use and inhabit. But reconfiguration of land tenure arrangements is prominently listed as an important aspect of achieving sustainability. The Forest Principles urge governments to establish "land tenure arrangements which serve as incentives for the sustainable management of forests."[169] Provisions of Agenda 21 speak of the need to provide secure land tenure, particularly for indigenous people and other local communities.[170] It specifically urges governments to protect the property rights of pastoral and nomadic groups,[171] and to ensure "effective land tenure" for rural people who live by farming, fishing, and forest harvesting.[172]

Taking an overview of these various provisions, we can see that the UNCED documents call on governments to take a series of steps all aimed, in one way or another, at giving local communities more influence over the management and exploitation of biological resources in their immediate environments. Governments should grant rights of participation in government decisionmaking to all local communities, implement a general policy of decentralizing administrative authority subject to central substantive standards, and take special measures to empower particular communities, based on their extreme vulnerability on the one hand, and their exceptional knowledge and other qualifications, on the other.

These provisions must be read together with the provisions recognizing global interests in the biological diversity of the same ecosystems. Local community empowerment has become a force for achieving the international goal of sustainability, and international oversight has become a means for ensuring the empowerment of local communities. A synergistic alliance of global and local interests that places state sovereignty in a new context has formed. The next

---

168. Agenda 21, *supra* note 2, para. 26.4.

169. Forest Principles, *supra* note 3, Principles/Elements, para. 5(a).

170. Agenda 21, *supra* note 2, para. 7.30(f).

171. *Id.* para. 12.28(c).

172. *Id.* paras. 32.15, 32.1 (definition of "farmer"), 32.69(b) (legal capacity of vulnerable groups). *See also id.* para. 3.8(f) (legal frameworks for land ownership), para. 11.12 (need to consider tenure patterns as part of forest rehabilitation programs), para. 14.17(b) (ensure equitable access to land, water and forest resources), para. 14.18(c) (assign clear titles, rights and responsibilities for land to encourage investment).

section discusses what some of the implications of this alliance may be.

III. "OWNERSHIP" IN AN ECOLOGICALLY INTERDEPENDENT WORLD: THE UNCED VISION OF PARTNERSHIPS

Who owns the rain forest? After UNCED, the formal answer to this question remains unchanged. The sovereignty of states over their natural resources has been reaffirmed and even strengthened. The authority of states to define and assign property rights through national law remains intact. Whether a forest or other ecosystem is held by a government as state property or by a landowner as private property, UNCED has brought no immediate changes to the titular designations of "sovereigns" and "owners."

Nevertheless, drawing on the "environmental" and "human rights" perspectives discussed in Part I, UNCED has advanced concepts of authority over natural resources that diverge from these seemingly straightforward notions of unitary control. Endorsing the imposition of new requirements on the management of natural resources, the UNCED documents seek to narrow the prerogatives of sovereigns and property owners and to ensure that biological resources sustainably serve the ecological and cultural interests of others who depend upon them in various ways.

States, in their exercise of sovereignty, are recast in an administrative role entailing duties to the international community on the one hand and to local communities within state boundaries on the other. To express the result in another way, both the international community and local communities are seen as having rights to resources formally held by sovereign states or persons entitled by them.[173] States in essence are asked to revise their legal systems according to the new international mandate.[174]

---

173. The UNCED documents thus follow the various ideas that have developed in American law characterizing the interests of non-owners in land, water, or other resources as a form of property or property-like expectations. Charles A. Reich, *The New Property*, 73 YALE L.J. 733 (1964); *see also* Glennon, *supra* note 31, at 33. Joseph Singer has characterized such recognition of rights of non-owners as the recognition of a "reliance interest" in property. Focusing on the plant-closing context, he argues that property law should recognize relationships that have evolved with the growth of an enterprise, and the mutual dependence of all who work to make an enterprise successful, as well as the extended network of community relationships that grow up around their endeavors. Workers and their communities, he contends, should be recognized as co-owners in decisionmaking about plant closures, and their rights to equitable relief or indemnification should be protected by the courts. Joseph William Singer, *The Reliance Interest in Property*, 40 STAN. L. REV. 611 (1988) [hereinafter Singer]. Emerging international recognition of

States' duties to the international community under the UNCED documents are, first, an extension of the concept of "planetary trust" discussed earlier.[175] As under the public trust doctrine in American law, the articulation of government duties is equivalent to a vesting of rights in the public at large.[176] The public in this case is not simply the citizens of a given state, but the international community in general.

At the same time, emerging state duties to protect local communities' access to and use of particular biological resources amount to a recognition of local communities as beneficial owners. Biological resources are acknowledged to have special meaning to particular communities, differentiated from their value to the public in general. In recognizing this special "connectedness" of local communities to their immediate environments, the UNCED documents import notions of customary local rights into the development of international environmental law.[177]

The UNCED documents' recognition of international and local interests in the resources formally held by sovereigns and current property owners reflects a conclusion that greater value will be realized by expanding the allocation of rights in biological resources, beyond the established definitions of title.[178] Implicit in these developments is a criticism of states' existing legal systems. Legal language of exclusive control has marginalized and excluded important local and global interests in sustaining biological diversity. The values of ecological and cultural relationships built around ecosystems have been left out of government programs to administer public lands, and omitted from the definitions of private property traded in markets. The UNCED documents seek to remedy this situation by

---

ecosystem rights for non-owners of biological resources amounts to adoption of comparable theories in the environmental protection context.

174. The reality of the UNCED mandate and the establishment of the array of rights and obligations envisioned in the UNCED documents will of course depend in large measure on the creation and structuring of the Sustainable Development Commission, the provision of adequate funding by states for the tasks set forth in Agenda 21, and the continued willingness of states to follow through on provisions that are non-binding.

175. *See supra* notes 31-34 and accompanying text.

176. Rose, *supra* note 45, at 711, 720-21, 727-30.

177. See *supra* text accompanying notes 59-86, regarding international human rights provisions addressing local rights in land. This recognition of local rights also resembles national laws delineating customary rights to land, based on a community's relationship with a particular place that makes that location uniquely valuable and essential for traditional community activities. See Rose, *supra* note 45, at 739-44, 758-60, regarding protection of the customary land uses of local communities under English precedent.

178. *See* Rose, *supra* note 45, at 760-61, 770. For analysis of reasons that markets and property rights as currently defined will not maximize value to society in such situations, see Glennon, *supra* note 31, at 5-6.

inducing states to revise their property regimes and their programs governing access to resources in order to reflect better the full array of interests at stake.

Central to this endeavor is a much-broadened concept of resource "use." Values of biological resources that previously went unrecognized are now incorporated in expanded definitions of the human "uses" of ecosystems needing protection.[179] Non-consumptive uses that are overlooked in existing legal systems receive attention in the UNCED documents, along with the consumptive uses of groups marginalized by dominant societies. In particular, the UNCED documents recognize the "services" that ecosystems provide to local communities (for example, as habitat for species that are harvested locally) and to the global community (for example, by providing international climate stabilization).[180] By calling on states to protect a broad range of ecosystem "uses," the UNCED documents in essence suggest that states should forsake old notions of unitary dominion over entire areas of land for new definitions of multiple ownerships tied to specific ecological functions and human dependencies.[181]

While they envision new and overlapping definitions of ownership, the UNCED documents are not an invitation to open access exploitation; to the contrary, they endorse clearer, more secure and better-enforced definitions of rights that correspond to existing ecological relationships and exclude incompatible exploitation. The goal is to eliminate violent conflicts over resources and bring about the internalization of external costs of resource use, through orderly transactions in which relevant values are more fully represented than they are in current markets and government decisionmaking.

The resulting vision of control over access to resources erodes notions of single ownership even as it seeks to exclude harmful exploitation through better-defined and better-enforced rights. Multiple global, national, regional and local interests have found expression in the UNCED documents in the various assertions about the rights, sovereignty, property, and roles of different players. Such overlapping definitions of authority make sense only if they are read as mutually-defining rather than conflicting, establishing relative authority within relationships of co-ownership, rather than authority of a single owner or sovereign acting without obligations to others.

The repeated references to the formation of new "partnerships" in the UNCED documents confirm this underlying shift away from

---

179. See *supra* note 10, regarding the expanded yet still-anthropocentric notions of value reflected in the UNCED documents.

180. *See supra* text accompanying notes 107-20, 154-71.

181. "The word 'use' itself, of course, is a traditional way to designate beneficial ownership in property held in trust." Rose, *supra* note 45, at 721 n. 38.

assignments of unitary control and toward a recognition of multiple centers of authority, coordinated through co-ownership arrangements tailored to particular locations and ecological circumstances.[182] As the term "partnership" suggests, the UNCED documents envision various entities contributing to joint endeavors and adjusting their activities in a coordinated fashion so that all may ultimately share in the benefits of collective enterprise involving pooled resources.[183] Within this general mandate, local communities, particularly those of indigenous people, are singled out as important "partners" regardless of whether they have already obtained any recognition of sovereign powers or property ownership within the states they inhabit.[184] Likewise, non-governmental organizations that voice international environmental concerns are among the recognized "partners" entitled to an authoritative role.[185]

The references to partnership imply reciprocal responsibilities as well as rights. The UNCED documents reject neither market mechanisms nor government administration of resources as possible tools for managing competing human demands on resources, but they implicitly call for tailoring the definitions and allocations of rights, including the rules affecting alienability of such rights, so as to reflect the values of all the various "uses" of those resources.[186]

---

182. References to "partnership" and "partners" appear no fewer than fifty-nine times in the various chapters of Agenda 21, *supra* note 2. The Rio Declaration, *supra* note 1, uses the same terms in the Preamble and Principles 7, 21, and 27.

183. THE AMERICAN HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE (William Morris ed., 2d ed. 1976) notes that the term "partner" "implies a relationship . . . in which each has equal status and a certain independence but also implicit or formal obligations to the other or others." The specialized biological definition of a "partner" offered in the OXFORD ENGLISH DICTIONARY (J.A. Simpson & E.S.C. Weiner eds., 2d ed. 1989) also suggests a pertinent ecological analogy: "Each of a pair or group of symbiotically associated organisms."

184. Notably, such partnerships are not seen as exclusively "private" or exclusively "public;" in calling for the formation of partnerships, the UNCED documents avoid sharp distinctions in status among different entities such as government agencies representing sovereign authority, private organizations acting through an exercise of property rights, and international institutions empowered through international agreements.

185. See *supra* notes 145, 153 regarding provisions that recognize authoritative roles for various groups and organizations in the UNCED documents. The roles of local communities and non-governmental organizations include access to the proceedings of the Sustainable Development Commission. *See supra* note 138; Agenda 21 paras. 38.5, 38.8, 38.11, 38.13. The expressed goal of recognizing and fostering the autonomy of self-defined groups including those that may be formed in informal or cultural ways without legal recognition from central governments, contrasts with a traditional hostility to the recognition of such entities under United States law. *See* Rose, *supra* note 45, at 742.

186. We might term the recommended modifications a "repackaging" of property rights to coincide with new understandings about how ecosystems function and how human beings depend upon them.

With their emphasis on the interdependence of such uses, the ideas of co-ownership that emerge in the UNCED documents necessarily involve restraints on the unilateral exploitation and alienation of biological resources.[187]

The ideas of interdependent co-ownership advanced in the UNCED documents may be concretely illustrated through some specific examples of resource management regimes. One system that embodies the concepts advanced at UNCED is Brazil's system of extractive reserves. In designated areas of state-owned forests, the government grants usufructuary rights to specific local communities to harvest rubber and other non-timber products, and it agrees to protect these communities from incompatible exploitation of resources by others.[188]

---

187. Arguments have been made that property should generally be assigned to unitary control, with narrow exceptions, given the transaction costs of negotiations among rights holders that must precede alienation of shared property to a third party. Richard A. Epstein, *Why Restrain Alienation?*, 85 COLUM. L. REV. 970, 971-72 (1985). Nevertheless, even this analysis of the costs of co-ownership acknowledges that restrictions on alienation may be necessary to deal with common pool problems, where "one person is not the exclusive owner of a single resource, but shares it in indefinite proportions with other claimants." *Id.* at 978. In essence, the UNCED documents suggest that biological resources such as forests are now known to have characteristics of common pool resources, given their multiple interdependent ecological functions, despite earlier beliefs that they could be readily severed into individually-alienable parcels. The trend toward recognition of multiple interests in the shared resources of biological diversity in the UNCED documents means that the envisioned structure of controls might well resemble systems of water law. As Eric Freyfogle has observed, "[i]f property law does develop like water law, it will increasingly exist as a collection of use-rights, rights defined in specific contexts and in terms of similar rights held by other people. Property use entitlements will be phrased in terms of responsibilities and accommodations rather than rights and autonomy." Eric T. Freyfogle, *Context and Accommodation in Modern Property Law*, 41 STAN. L. REV. 1529, 1531 (1989). More generally, the reasons for abandoning the language of absolute rights and acknowledging the responsibilities that accompany rights have been addressed in MARY ANN GLENDON, RIGHTS TALK: THE IMPOVERISHMENT OF POLITICAL DISCOURSE (1992). "What we need therefore is not a new portfolio of 'group rights,' but a fuller concept of human personhood and a more ecological way of thinking about social policy." *Id.* at 137. *Compare* Williams, *supra* note 65 at 660-64 (emphasizing the usefulness of rights discourse to indigenous peoples confronting the inequities of existing legal regimes.)

188. For descriptions of extractive reserves and commentary on their potential effectiveness in achieving sustainable development of resources, see Mary Helena Allegretti, *Extractive Reserves: An Alternative for Reconciling Development and Environmental Conservation in Amazonia* [hereinafter Allegretti] *in* ALTERNATIVES TO DEFORESTATION: STEPS TOWARD SUSTAINABLE USE OF THE AMAZON RAIN FOREST (A.B. Anderson ed., 1990); Stephan Schwartzman, *Extractive Reserves: The Rubber Tappers' Strategy for Sustainable Use of the Amazon Rainforest* [hereinafter *Extractive Reserves*] *in* FRAGILE LANDS OF LATIN AMERICA: STRATEGIES FOR SUSTAINABLE DEVELOPMENT (John O. Browder ed., 1989); Stephan Schwartzman, *Land Distribution and the Social Costs of Frontier Development in Brazil: Social and Historical Context of Extractive Reserves*, [hereinafter *Social Costs*] *in* ADVANCES IN ECONOMIC BOTANY (forthcoming, 1992) (on file with *Tennessee Law Review*); Stephan Schwartz-

Each local community holds its rights to the renewable resources of the forest subject to a requirement that its uses must be "sustainable."[189] Although it may freely market the products obtained from the forest (such as rubber or Brazil nuts) the community has no power to sell the land itself, which remains in the hands of the government.

Brazil's administrative and legislative measures creating extractive reserves were driven forward by a powerful alliance formed in the 1980's between the National Council of Rubber Tappers, led by Chico Mendes, and Washington, D.C.-based environmentalists. The dispossession of rubber tapper communities from their traditional territories went hand-in-hand with the massively destructive burning and clearing of the rain forest for ranching and agriculture. The two groups found common goals, merging their respective crusades for local land rights and global rain forest protection. The leverage that each group gained with the Brazilian authorities as a result of the coalition has been vividly depicted in the writings of social scientists, some of whom were directly involved in the negotiations.[190] The alliance initially circumvented the legal avenues available under a recalcitrant Brazilian government, resorting to pressure in the United States Congress aimed at reducing World Bank funding for Brazilian development projects.[191] Nevertheless, it was the Brazilian government itself that ultimately provided the legal mechanisms necessary for extinguishing competing land claims, granting clear use rights to communities of rubber tappers, and protecting those rights from

---

man, *Deforestation and Popular Resistance in Acre: from Local Social Movement to Global Network*, The Centennial Review, Spring 1991, at 397-422 (on file with *Tennessee Law Review*).

189. *See* Decree No. 98.897 of Jan. 30, 1990, Diario Oficial 2122-23 (1990) (defining extractive reserves and establishing sustainable use conditions to be implemented through contracts between the resource users and the government) (on file with *Tennessee Law Review*).

190. *See* Margaret E. Keck, The Acre Rubber Tappers' Movement and Environmental Politics in Brazil: An Interpretation (paper prepared for presentation at the 1991 Annual Meeting of the American Political Science Association, Washington, D.C., August 29 - September 1, 1991) (on file with *Tennessee Law Review*); *see also* Allegretti, *supra* note 188; *Extractive Reserves*, *supra* note 188; *Social Costs*, *supra* note 188.

191. The events are recounted in ANDREW REVKIN, THE BURNING SEASON: THE MURDER OF CHICO MENDES AND THE FIGHT FOR THE AMAZON RAIN FOREST 185-230 (1990). While the rubber tappers amended their assertion of rights to encompass an environmental trusteeship of the rain forest, see Platform of the National Rubber Tappers Council, *supra* note 89, the idea of international rights to Brazil's rain forest gained momentum. Marking the occasion of Chico Mendes' assassination, the Congressional Record contains the following remarks by Senator Kasten: "America cannot be true to itself if it sanctions the destruction of the world's common heritage. . . . Because [Brazil's rainforests] perform a vital service to the whole global environment . . . they cannot properly be called simply 'Brazilian

encroachment through government ownership and enforcement.[192]

The story of the first Brazilian extractive reserves thus presaged and exemplified the union of environmental and human rights concerns that found expression in the UNCED documents, as well as the reliance on state legal mechanisms for implementation. A paradigm for subsequent efforts, the extractive reserves regime serves the interests of the international community in the preservation of an intact rain forest, as well as the interests of local communities in their means of subsistence and way of life, by means of a legal structure administered by the state. While the economic viability of extractive reserves has yet to be demonstrated, advocates note that their value must be calculated not merely in terms of monetary return to the forest inhabitants from harvested products, but also in terms of global and regional benefits that warrant continuing support of the system by a wider community.[193]

Brazilian extractive reserves are but one example of "co-management" systems for the administration of natural resources that could serve the international mandate of the UNCED documents for the preservation of biological diversity. Such systems generically involve "the sharing of power and responsibility between the government and local resource users."[194] The Global Biodiversity Strategy issued shortly before UNCED by the United Nations Environment Programme, the World Resources Institute, and the World Conservation Union explicitly endorses co-management regimes including extractive reserves, not merely as means for serving local subsistence needs, but also as viable methods for achieving international goals of biodiversity conservation.[195] A variety of co-management arrangements already exist that respond to the UNCED mandate, providing empowerment of local communities in the use and management of living resources,

---

rainforests.' The fact is, we need them and we use them—so they're our rainforests too." 135 CONG. REC. 817 (daily ed., Jan. 3, 1989) (statement of Sen. Kasten). The formation of such partnerships between U.S. environmental groups and local organizations from other countries is addressed in Wirth, *supra* note 7.

192. For an account of the initial laws and decrees establishing extractive reserves, see Peter C. L. Roth, *The Emerging Role of the Extractive Reserve in the Enforcement of Brazilian Deforestation Controls*, 2 COLO. J. INT'L ENV. L. & POL'Y 247, 270-74 (1991).

193. See *Social Costs*, *supra* note 188 for an articulation of these arguments.

194. Fikret Berkes et al., *Co-management: The Evolution in Theory and Practice of the Joint Administration of Living Resources*, ALTERNATIVES, September/October 1991, at 12. These writers note that the term may cover a broad variety of arrangements that differ in the amount of control granted to the local community: "[T]here is a continuum of co-management arrangements from those that merely involve, for example, some local participation in government research being carried out, to those in which the local community holds all the management power and responsibility." *Id.* (footnote omitted).

195. GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 84-87.

within a broader framework oriented toward attainment of national and international goals.[196]

While the term "co-management" is generally used to describe bilateral arrangements between national governments and local communities, similar notions of co-ownership have been extended, in different terminology, to arrangements among multiple government bodies and private actors who join together using the various governmental and ownership powers at their disposal, in order to protect and sustainably use resources that are ecologically related, though separately owned. "Biosphere reserves," a central focus of the Man and the Biosphere Program of the United Nations Education, Scientific, and Cultural Organization (UNESCO), have provided a starting point for even more far-reaching proposals for "bioregional management."[197] A model biosphere reserve involves a protected "core area" surrounded by a "buffer zone" and a "transition area" in which human activities are variously restricted through government ownership and environmental and zoning regulations.[198] More broadly, bioregional management concepts entail an interactive mosaic of multiple geographic areas in a variety of ownerships, providing a range of habitats and supporting different human uses, but forming an ecologically and culturally related whole.[199] The different actors are brought together in a participatory process through any of a variety of institutional mechanisms, for the purpose of jointly planning and managing their collective uses of the ecosystem of which they are a part.

These brief examples of co-ownership concepts in ecosystem management are sketched here for the purpose of highlighting some common themes that are implicitly endorsed by the UNCED documents. All of these models entail recognition of multiple authoritative entities and empowerment of marginalized groups. Yet at the same time, these ideas involve efforts to achieve newly-unified forms of human organization, growing out of agreements and joint action by the various autonomous parties, bridging entrenched distinctions between "public" and "private" entities and among "international," "national" and "local" functions.

---

196. *See, e.g., id.* at 85-87 (summarizing initiatives in, *inter alia*, the Philippines and northern Canada); David S. Case, *Subsistence and Self-Determination: Can Alaska Natives Have a More "Effective Voice"?*, 60 U. COLO. L. REV. 1009 (1989) (discussing the successes and failures of various laws affecting the roles of Alaska Natives and governments in wildlife management in Alaska); Gail Osherenko, *Can Comanagement Save Arctic Wildlife?*, ENVIRONMENT, July/August 1988 at 7-33 (analyzing wildlife management systems in Alaska and the Canadian North).

197. GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 97-104.

198. *Id.* at 100.

199. *Id.* at 97-100. A comprehensive description of bioregional principles may be found in KIRKPATRICK SALE, DWELLERS IN THE LAND: THE BIOREGIONAL VISION (1985).

In summary, the UNCED documents call for a newly pluralistic international order for the purpose of fostering biological diversity and ensuring the renewability of living resources. Forsaking rigidly hierarchical organization and pre-existing boundaries in resource management regimes, the UNCED documents envision new centers of autonomy and new forms of interconnection among them. The image conveyed is of a dynamic, interactive, non-linear process for decisionmaking that fosters access by non-governmental organizations to international institutions and international networking of individuals and groups, in ways that may bypass state bureaucracies. But the empowerment of local communities and other entities is not intended to create disagreement or conflict. To the contrary, the UNCED documents demand a process that, while newly complex, is unified, holistic, integrated, cooperative, and peaceful. Agenda 21 refers repeatedly to "mechanisms" for ensuring that international goals of sustainable development are achieved.[200]

Concealed in this hopeful vision of a unified and smoothly functioning international regime[201] are tensions between concepts of local empowerment and notions of global unification. What is the meaning of local community autonomy in an ecologically interdependent global order? The alliance at UNCED between the "environmental" perspective and the "human rights" perspective is the manifestation of convergent but not necessarily identical views on this question.

From the paradigmatic "environmental" perspective, the autonomy and rights to self-management of local communities are important means for achieving global ecological goals. Shifting power to local communities is a way of fostering biological diversity by placing resources in the hands of people who are particularly knowledgeable about, and interested in, the preservation of their environment, and by taking control away from those who have profited, at the expense of local and global interests, from over-exploitation. This perspective takes a predominantly instrumental view of local communities: local rights are means to global ends. The advocated delegation of authority to local communities is based on a perception of local capabilities and qualifications in achieving those goals.[202]

While modern environmentalists reject any attempt to confine indigenous peoples or other groups to a "museum" or "zoo,"

---

200. *E.g.*, Agenda 21, paras. 3.4, 3.7(c), 3.8, 7.20(f), 7.35, 8.3(c), 10.5, 12.28(c), 12.44, 12.57(b), 15.5(e), 17.6, 18.54, 18.90, 27.5-27.11.

201. See HURST HANNUM, *supra* note 61, for a similarly hopeful view of the possibilities for encompassing "autonomous" entities within a unified international order. Nathaniel Berman raises the quite different possibilities of irreconcilable tensions among self-defining groups, in his review of the book. Nathaniel Berman, Book Review, 85 AM. J. INT'L L. 730 (1991).

202. *See supra* notes 35-49 and accompanying text.

acknowledging the benefits to all of allowing communities to continue to evolve with, and innovate in, their environments, an "environmental" perspective nevertheless values local cultures for the values that they hold, the forms of social organization that they develop, and the traditional wisdom that they have.[203] In short, the "environmental" perspective has an intense interest in the identity of the local community, and it implicitly reserves for the international community a continuing right to judge whether that local identity coincides with the environmental agenda, justifying a continuing designation of the local community as an authorized manager of resources of global importance.[204]

The rights of the international community in the continued viability of biologically diverse ecosystems take the form of a property-like interest in the entire structure of international, national and local rights and duties—ranging from large-scale definitions of sovereign authority and duties respecting protection of global resources to small-scale norms for allocating and preserving resources among members of specific local communities.

As Professor Carol Rose has observed, the interests of the general public in the allocation of specific rights and duties to individuals and groups for the sake of producing broad social benefits may be termed "meta-property."[205] This overarching public interest encompasses all the various forms of national and local property regimes that affect the allocation of local access to resources and the ensuing benefits provided to the international community as a whole. The rights of local communities, if formulated to serve global interests, are themselves an entitlement of the international community. By extension, local communities are not simply resource users but also trustees who carry out, through delegated authority, the trust responsibilities of national governments to a wider community. Environmentalists in essence seek to scrutinize the performance of these trusteeship responsibilities.

A paradigmatic "human rights" perspective, by contrast, emphasizes local rights to autonomy and self-management, and views the

---

203. For a straightforward statement of the difficulties conservationists have in readily accepting the innovations of local communities, when the changes may lead to increased harvesting of resources, see Clad, *supra* note 49, at 50-51.

204. *E.g.*, GLOBAL BIODIVERSITY STRATEGY, *supra* note 9, at 79-80 (advocating local control of resources, but emphasizing that "governments should enforce basic norms of stewardship on behalf of the wider society and future generations," and noting the inability of some local communities to carry out the goals of the broader community). *See also* Zygmunt J.B. Plater, *Keynote Essay: A Modern Political Tribalism in Natural Resources Management*, 11 PUB. LAND L. REV. 1 (1990) [hereinafter Plater] (an environmentalist's assessment of claims by local communities).

205. Rose, *supra* note 45, at 747 (citing remarks of J. Krier).

environment as closely affiliated with local community identity.[206] For indigenous peoples, the issue of self-determination has been especially prominent, but similar questions about whether resource decisions should be assessed locally or centrally arise in other contexts as well. Disagreements between environmentalists and local communities over harvesting rights, whether they involve fishing, whaling, logging, or access to subsurface resources, reflect differing views about the definition and evolution of the authoritative "community."[207]

The UNCED documents offer no fixed prescription for a continuing alliance of global and local views on the management of biological diversity, and the human place within ecosystems. Multiple "owners" have been posited, and a dialogue among them has begun, but the full meaning of autonomy and interdependence, and of local community and global community, has yet to be negotiated.

206. *See supra* notes 59-87 and accompanying text.

207. Plater, *supra* note 204 (discussing disagreements over resource use between environmentalists and local communities in the United States); Wirth, *supra* note 7, at 2660-61 (noting potential conflicts of interest between U.S. environmental groups and organizations representing local interests in other countries); Reiser, *supra* note 33, at 410 (noting potential incompatibility between community rights of access to resources and public rights to ecological preservation).

"Environmental Security" in the United Nations: Not a Matter for the Security Council

CATHERINE TINKER*

At an historic summit on January 31, 1992, the heads of state and government of the fifteen members of the United Nations Security Council concluded in their final declaration that:

> The absence of war and military conflicts amongst States does not in itself ensure international peace and security. *The non-military sources of instability in the* economic, social, humanitarian and *ecological fields have become threats to peace and security.* The United Nations membership as a whole needs to give the highest priority to the solution of these matters.[1]

This quite remarkable and far-reaching agenda from the leaders of the Security Council does not necessarily require Security Council action on each of the "non-military sources of instability" if other organs of the United Nations, such as the General Assembly or the Economic and Social Council (ECOSOC) can better address them. But the agenda raises the distinct possibility of a reinterpretation of the Security Council's mandate beyond the scope of traditional military definitions of international peace and security.

At the time of the Security Council summit meeting, some North American non-governmental organizations urged that the United Nations Security Council place environmental matters on its agenda.[2]

---

* Visiting Associate Professor of Law, The State University of New York at Buffalo. LL.M. 1989, New York University; J.D. 1978, George Washington University; M.A. 1973, Occidental College; B.A. 1971, St. Olaf College. The author thanks Dr. Paul Szasz for his helpful comments on an earlier draft of this Article.

1. Note by the President of the Security Council, U.N. SCOR, 3046th mtg. at 3, U.N. Doc. S/23500, (1992) (emphasis added). This paragraph was the subject of a discussion convened by the author on May 11, 1992, at the U.N.A.-U.S.A. in New York City with United Nations Secretariat officials, academics, permanent representatives of member states to the United Nations, and representatives of United Nations specialized agencies and non-governmental organizations. The discussion was chaired by Professor Ruth Wedgewood of the Yale Law School.

2. The National Audubon Society, the Environmental Defense Fund, and the Natural Resources Defense Council (NRDC) held a press conference in New York on January 29, 1992, to urge the Security Council summit meeting to put the environment at the top of its agenda.

In its press release the NRDC claimed that "the concept of 'international security' has changed literally overnight—from dominance by East-West Superpower relations to a far broader notion that should include 'planetary security' from potentially irreversible environmental deterioration." (copy on file with author).

In evaluating the soundness of this recommendation, it is important to note that environment and development issues are already on the active agenda of the General Assembly. The General Assembly decided in 1989 to convene the United Nations Conference on Environment and Development,[3] and the General Assembly established the Intergovernmental Negotiating Committee on the draft climate change convention.[4]

Both the Security Council and the General Assembly have powers under the United Nations Charter to recommend action to preserve international peace and security.[5] Thus, either organ may address environmental hazards if such hazards threaten peace. There may be some risk to world stability if the Security Council treats environmental issues, such as conflicts over resource allocation or pollution, as direct threats to international peace and security. That risk could easily develop if the United Nations Security Council decides to take collective security action against a nation because of an environmental concern, such as the building of a dam that could create refugee or flooding problems across the border. The utility of collective security action is questionable in this context, since much environmental damage is irreversible. Preventive action more suitable to environment and development concerns does not fit under Chapter VII of the Charter, which was designed to stop classic military transborder invasions and the like.

The concept of "environmental security" springs from concerns raised during the Stockholm Conference on the Human Environment in 1972.[6] In a 1989 article, Jessica Tuchman Mathews raised the possibility of "redefining security" to include environmental harm as a form of aggression.[7] Eerily, just one year later, environmental damage was used as a weapon in the Gulf War. As bombed refineries began to burn and broken pipes spilled oil into the Gulf, memorable media coverage revealed pictures of birds caught in oil slicks, and landscapes obscured by thick haze heavy with dark particles. The idea of environmental security took on fresh meaning, particularly

---

3. G.A. Res. 44/228, U.N. GAOR, 44th Sess., Supp. No. 49, at 151, U.N. Doc. A/44/49 (1990).

4. G.A. Res. 43/53, U.N. GAOR, 43d Sess., Supp. No. 49, at 133, U.N. Doc. A/43/49 (1989). The Climate Change Convention was opened for signature in June 1992, during UNCED in Rio de Janeiro, Brazil.

5. U.N. CHARTER art. 39 (Security Council); U.N. CHARTER art. 11 (General Assembly).

6. Declaration on the Human Environment, *reprinted in* 11 INT'L LEGAL MATERIALS 1416 (1972) (U.N. A/Conf. 48/14).

7. Jessica Tuchman Mathews, *Redefining Security*, FOREIGN AFF., Spring 1989. This article was later expanded into a book titled PRESERVING THE GLOBAL ENVIRONMENT: THE CHALLENGE OF SHARED LEADERSHIP (Jessica Tuchman Mathews ed., 1991).

when reports of strange climactic alterations beyond Kuwait's boundaries arrived from India and cyclones flooded Bangladesh.

In 1991, the Security Council passed a resolution that included the first determination under international law of a state's liability for harm to the environment itself, apart from direct injury to people or property, and for the depletion of natural resources. Within the limited context of the ceasefire with Iraq, the defeated belligerent was declared "liable under international law for any direct loss, damage, *including environmental damage and the depletion of natural resources*, or injury to foreign Governments, nationals and corporations, as a result of Iraq's unlawful invasion and occupation of Kuwait."[8] Because of the unusual circumstances surrounding this declaration, its precedential value may be limited. Further, the problems associated with establishing a claims settlement mechanism[9] and a theory of standing to claim on behalf of the environment or the resources[10] have not been resolved. Nevertheless, this Security Council resolution indicates a willingness to recognize harm to the environment in the context of a Chapter VII action,[11] although the environmental damage referenced in the resolution occurred as the result of armed conflict. It was not the original act of aggression that triggered the Chapter VII response from the Security Council.

---

8. S.C. Res. 687, at ¶ 16, *reprinted in* 30 INT'L LEGAL MATERIALS 846 (1991) (emphasis added).

9. No money damages have been paid to date as a consequence of Iraq's continuing refusal to agree to the United Nations plan for lifting the embargo on sales of Iraqi oil on condition that the proceeds be administered by the United Nations. Under this plan, the proceeds from any sales would first go to cover United Nations expenses in connection with the situation in Iraq and occupied Kuwait and the aftermath of the war. Then funds would go into a claims settlement fund, with a certain portion, but less than half of the proceeds, going to Iraq to provide food, medicine, and other necessities for its citizens. Without any funds to satisfy claims, there has been no hurry to begin processing claims of damage.

10. An ombudsperson or "Friend of the Planet" could come forward on behalf of the environment and the earth's natural resources to claim money damages from Iraq and to apply any damages awarded and recovered to environmental clean-up and restoration efforts.

11. Collective action under Chapter VII is permissible only when the Security Council determines there is a threat to international peace and security, a breach of the peace, or an act of aggression. This threshold determination, made under Article 39 of the Charter, is necessary to "open the door" to further collective action on an escalating scale, by applying each successive article of Chapter VII. The Security Council can investigate any dispute or situation, recommend methods of adjusting disputes, apply economic sanctions of various kinds, and, failing all else, take military action against an aggressor pursuant to Articles 41 and 42. Chapter VII only has been used twice in the nearly fifty years of the United Nations' existence, once in Korea in the 1950s and once in the Gulf in the 1990s. Both times, one nation's army invaded another nation across a national border, an act of aggression clearly affecting international peace and security.

This Article will evaluate the practical consequences of environmental security. Can a nation's depletion of resources or activities that harm the environment—acts that are not currently prohibited by international law—justify a Security Council response? If threats of or actual damage to the global environment or destructive development practices by a nation are themselves "threats to international peace and security, breaches of the peace, or act[s] of aggression"[12] within the meaning of the United Nations Charter, the Security Council may act to coerce a nation to discontinue the harmful practices. Will triggering Security Council action under Chapter VII accomplish this goal? Will it be too late to stop or reverse environmental damage?

Part I of the Article first examines how the Security Council becomes involved in a conflict and then reviews the options for collective security action by the Security Council, beginning with peaceful means of settlement and escalating to the imposition of economic sanctions and finally to the use of military force. Part II determines whether the definition of a threat to international peace and security or an act of aggression under Chapter VII of the United Nations Charter is applicable to environmentally harmful acts. The tension between sovereignty and intervention in the internal affairs of a state in the name of environmental protection or sustainable development is explored. The utility of sanctions or military force against nations responsible for environmental harm is considered.

Part III considers other alternatives within the United Nations system, including the role of General Assembly and ECOSOC, for addressing incidents of serious environmental damage. Part IV discusses the potential of the International Court of Justice and arbitral panels to resolve disputes over scarce resources or environmental damage. In cases of environmental threats, these alternatives to Security Council coercive action are more appropriate to achieve the goal of prevention of environmental degradation or disaster and exhaustion of planetary resources. Environmental, social, or economic affairs often involve conflicting values and the need to choose between two alternatives that offer different benefits, costs, and allocation of risks over time. It is important that the judicial, arbitral, or institutional mechanisms for making this choice be as truly representative as possible.

This Article concludes that environmental damage may be a "threat to international peace and security," but should not be on the agenda of the Security Council as it is presently constituted. Otherwise, the votes of the five World War II victors—with the veto power—may be used to justify intervention into a nation's affairs without its consent under Chapter VII of the United Nations Charter.

---

12. U.N. CHARTER art. 39.

Unless and until the membership of the Security Council changes, the General Assembly and ECOSOC should continue to debate environment and development issues. International action as needed should be initiated by the General Assembly. Such a course promotes preventive measures, sustainable development, and the peaceful settlement of disputes.

I. THE SECURITY COUNCIL

The Security Council has primary responsibility—but not exclusive responsibility—within the United Nations system for the maintenance of international peace and security under Article 24(1).[13] Only the Security Council has the authority to make decisions that member states are obligated to carry out under Article 25 of the Charter.[14] The Security Council, which is of workable size and high prestige, deals with the most serious global issues. Immediately following World War II and throughout the Cold War, global issues were seen in military terms rather than in social or economic terms. Today the question of "security" cannot be limited in this way. The gravest threats to stability and the continuity of the world now may come from economic and social problems: dislocations, refugees and migrants, poverty, population growth, scarcity of resources necessary to sustain life, and other similar problems. These devastating social and economic conditions, in and of themselves, are threats to international peace and security, and *cause* certain actions by nations that are clearly acts of aggression and breaches of the peace in the more traditional military sense.

How does the Security Council become involved in disputes? As in the case of Korea, the Secretary-General may bring any matter "which in his opinion may threaten the maintenance of international peace and security" to the attention of the Security Council pursuant to Article 99.[15] Both member states and non-members of the United Nations can bring matters to the attention of either the Security Council or the General Assembly pursuant to Article 35.[16] The Security Council may discuss an issue on its own or upon recommendations from the General Assembly pursuant to Articles 10 and 11(2) and (3).[17] If the Security Council is "exercising its functions" under the Charter to resolve a dispute or situation, the General Assembly cannot make any recommendations unless requested to do

13. U.N. CHARTER art. 24, ¶ 1.
14. U.N. CHARTER art. 25.
15. U.N. CHARTER art. 99.
16. U.N. CHARTER art. 35, ¶ 1. Article 35 is not limited to peace and security issues.
17. U.N. CHARTER arts. 10, 11, ¶¶ 2, 3.

so by the Security Council.[18] The Secretary-General must notify the General Assembly of any matters relative to the maintenance of international peace and security being handled by the Security Council. If the Security Council ceases to deal with such matters, the Secretary-General must immediately notify the General Assembly.[19] A state which is not a member of the Security Council may participate in discussions without a vote if the Council decides that the interests of the state are directly affected or if the state is a party to a dispute being considered by the Council.[20]

The Security Council may attempt to resolve disputes by peaceful means under Chapter VI at any time. If there is an Article 39 determination that a threat to international peace and security, a breach of the peace, or an act of aggression exists, the Security Council must set forth principles for a peaceful settlement or conduct an investigation before taking coercive action.[21] The Council may request the Secretary-General to appoint a special representative or use good offices to resolve the dispute.[22] Failing all else, the Security Council may resort to economic sanctions and ultimately the use of force.[23] Each progressive application of enforcement measures under Chapter VII depends on the initial finding of a threat to, or a breach of, international peace and security or an act of aggression.[24]

In the Charter scheme, any military action authorized by the Security Council is under the control and direction of the United Nations through the Military Staff Committee with United Nations troops committed by member states under Article 43 agreements.[25] In Korea and in Kuwait, however, the troops were under the military command of United States Army officers. The failure of the international community to complete the design envisioned under Chapter VII regarding international control of United Nations forces through a functioning Military Staff Committee contributes to the fear of less-powerful countries that forces under the command of a single state or a limited "coalition" will be used against them. Many states resist expansion of the Security Council's agenda or object to attempts to broaden the list of subjects that relate to the definition of international peace and security on this basis.

II. Definition of International Peace and Security

The two prior occasions in which the Security Council found a "threat to international peace and security, breach of the peace, or

18. U.N. Charter art. 12, ¶ 1.
19. U.N. Charter art. 12, ¶ 2.
20. U.N. Charter arts. 31-32.
21. U.N. Charter arts. 33-34.
22. Thomas M. Franck, Nation Against Nation: What Happened to the U.N. Dream and What the U.S. Can Do About It 134-35 (1985).
23. U.N. Charter arts. 41-42.
24. U.N. Charter art. 39.
25. U.N. Charter art. 43.

an act of aggression" each involved armies crossing national borders invading other nations.[26] Except for enforcement actions under Chapter VII of the Charter, intervention by the United Nations in the internal affairs of a state is prohibited.[27]

To permit United Nations intervention in the name of sustainable development, "environmental security" must be part of international peace and security. Enforcement action if the safety or stability of the planet is endangered by unsustainable development might be based on one of two theories: (1) a human rights theory of fundamental rights to sustainable development or a healthy environment including clean air, water, or soil;[28] and (2) a planetary trust[29] or common heritage theory that places natural resources beyond the territorial control of any single state. Under the first theory, a type of "humanitarian intervention" or "humanitarian assistance" similar to relief actions to aid the Kurds in Iraq after the Gulf War could be used to assist people trying to prevent environmental disaster or during an environmental crisis with or without the consent of the government involved. Under the second theory, a nation does not have absolute sovereignty over natural resources within its territorial boundaries, but is only holding them in trust for the planet as a common resource. Therefore, if the actions of a nation endanger the resources or the habitability of the planet, international action is justified to protect the resources. Neither theory, however, has gained full acceptance in international law. Another source of prohibition against environmentally harmful acts in international environment law is Principle 21 of the Stockholm Declaration,[30] which balances the tension between rights and responsibilities.

Environmental problems are directly related to the welfare and survival of citizens.[31] Linked to economics, "environmental security" may be understandable as "part of a larger, broader definition of what security really means."[32] One cause of military conflict in the

---

26. *See supra* note 11.

27. U.N. CHARTER art. 2, ¶ 7.

28. W. PAUL GORMLEY, HUMAN RIGHTS AND ENVIRONMENT: THE NEED FOR INTERNATIONAL COOPERATION (1976). *See also* The Rio Declaration on Environment and Development, Principle 3, *reprinted in* 31 INT'L LEGAL MATERIALS 874 (1992) (U.N. Doc. A/Conf.151/5/Rev. 1).

29. EDITH BROWN WEISS, IN FAIRNESS TO FUTURE GENERATIONS: INTERNATIONAL LAW, COMMON PATRIMONY, AND INTERGENERATIONAL EQUITY (1989). Planetary trust theory consists of the belief that one generation receives the planet from a previous generation to hold in trust for the following generation. *Id.* at 2.

30. Declaration on the Human Environment, principle 21, *reprinted in* 11 INT'L LEGAL MATERIALS 1416, 1420 (1972) (U.N. Doc. A/Conf.48/14). See *infra* note 55 and accompanying text (for text of Principle 21). Principle 21 is repeated verbatim as a binding legal principle in the new United Nations Convention on Biological Diversity and in slightly altered form in the Rio Declaration.

31. ENVIRONMENTAL PROBLEMS: A GLOBAL THREAT at 16 (Report of the Twenty-Fourth United Nations of the Next Decade Conference 1989).

32. *Id.*

future is likely to be environmentally based, stemming from depleted resources, environmental refugees, and transboundary pollution. Furthermore, some environmental damage is permanent and irreversible, requiring prompt, high-priority action to prevent such damage.

The implication of calling environmental hazards threats to peace and security is to affect the type of response elicited. If using the security concept helps publicize the issues to governments and the general public, whose consumer habits and national priorities may change accordingly, then it is useful. On the other hand, if it means that traditional coercive responses will be mobilized against a nation under Chapter VII the results may be counter-productive or not very useful, particularly if damage to the environment has already occurred. For example, sending in military troops under United Nations auspices to prevent trees being cut down or to stop the building of a factory using polluting technology is clearly inappropriate and may itself be a threat to international peace and security. The single factory may be relatively harmless, but the cumulative effect of many such factories causes great environmental damage. Use of force to stop one nation's development activities—even for the goal of preserving natural resources—is not a use of the United Nations police power contemplated by the drafters of the United Nations Charter. Yet such action could be the logical consequence of "redefining security" to include environmental degradation or resource depletion.

Economic sanctions, imposed by either the Security Council or the General Assembly, against a global polluter nation may be attractive as symbolic action by the international community. But such sanctions ultimately punish the citizens of the target state more than its leaders, citizens who may already be suffering from the direct effect of the environmental hazard. Sanctions are also of limited effectiveness when used in isolation from other diplomatic or military actions.[33] Further study is needed to determine if economic sanctions are more or less likely to make a state alter its domestic policy, particularly if the environmentally harmful policy serves some locally useful purpose such as increasing the gross national product of the state or raising the standard of living of some citizens.

If environmental damage has already occurred, as in a large-scale environmental disaster, then neither sanctions nor armed force will be particularly useful, regardless of whether the government was negligent or acted intentionally in allowing the disaster to occur. In that situation, the most appropriate response from the United Nations is to bring in relief workers to aid the victims and to clean up the damage, seeking the cooperation of the government involved. This is a far cry from the type of collective security action provided by Chapter VII of the United Nations Charter, which was designed to deter a recalcitrant nation from escalating aggression or threats into

33. *See* MICHAEL P. MALLOY, ECONOMIC SANCTIONS AND U.S. TRADE (1990).

further military action. For example, as a means of evicting Iraq from Kuwait after Iraq's armed occupation of its neighboring sovereign state, Chapter VII offered an ideal sequence of actions. The Security Council remained firm in its position and tried to postpone a battle by creating opportunities for Iraqi withdrawal short of armed conflict. When these efforts failed, the Security Council used force.

It is difficult to imagine how this same set of increasingly coercive measures imposed by the Security Council would deter a nation from exploiting its resources or manufacturing desired goods in environmentally harmful, but currently legal, ways if the economic rewards were high enough, even if the gains would be short-term. More effective in encouraging nations to pursue policies of sustainable development[34] is the development of strong international environmental law, including liability provisions, global institutions to implement economic incentives not to pollute (such as carbon taxes or tradable permits calculated on a global basis for emission controls on harmful gases), and methods of national accounting (which reflect the high cost of resource depletion or reckless development). A strong body of international environmental law needs to articulate specific duties, responsibilities, and rights in this area. Developed nations must contribute the financing and transfer of technology to make global sustainable development a reality.

Aside from doubts about the efficacy of collective security measures in cases of environmental degradation, hazard, or resource depletion, there is another risk in adding environmental security to the Security Council agenda. A nation may act in self-defense or through a regional organization or alliance in collective self-defense in the absence of Security Council action. Article 51 acknowledges that "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations until the Security Council has taken measures necessary to maintain international peace and security."[35] It is not too big a leap to call severe transboundary pollution the equivalent of an armed attack. Similarly, regional organizations are recognized under Chapter VIII, Article 52(1):

> Nothing in the present Charter precludes the existence of regional arrangements or agencies for dealing with such matters relating to the maintenance of international peace and security as are appropriate for regional action, provided that such arrangements or

---

34. The term "sustainable development" was used in the Brundtland report prepared by the World Commission on Environment and Development, OUR COMMON FUTURE, 1987, to refer to development that meets both present and future environment and development objectives.

35. U.N. CHARTER art. 51.

> agencies and their activities are consistent with the Purposes and Principles of the United Nations.[36]

A broad definition of international peace and security thus opens the door to the possibility of unilateral or regional military response to environmental activity, and to collective security action through the Security Council:

> In the event of a serious threat to its environment security, a state might be eager to legitimize unilateral actions—such as attacking a tanker which is about to pollute the sea or bombing a highly polluting factory. . . . While on the one hand including issues related to environmental security can generate concern at the top level for the threats to the global environment, on the other hand it may unintentionally lead to an extension of the unilateral use of force.[37]

The characterization of environmental harm as a threat to international peace and security is not necessary for peaceful settlement of disputes. The expansion of the term "international peace and security" may even lead to mischief if it is used to trigger coercive United Nations or unilateral action against a nation based on that nation's own domestic policy choices on development and environmental protection without a clear mandate under international law and the United Nations Charter.

III. THE GENERAL ASSEMBLY RATHER THAN THE SECURITY COUNCIL IS THE APPROPRIATE UNITED NATIONS BODY FOR ENVIRONMENT AND DEVELOPMENT ISSUES

The General Assembly is composed of representatives of every member state of the United Nations with each state exercising one vote.[38] A two-thirds majority of the members is necessary to approve "recommendations with respect to the maintenance of international peace and security" under Article 18.[39] The General Assembly is empowered to discuss and make recommendations on the "general principles of cooperation in the maintenance of international peace and security" to the member states, to the Security Council, or to both.[40] Unless the Security Council is currently seized of an issue, the "General Assembly may recommend measures for the peaceful

---

36. U.N. CHARTER art. 52, ¶ 1.

37. Nico Schrijver, *International Organization for Environmental Security*, 20(2) BULL. OF PEACE PROPOSALS 115, 116 (1989). The author cites to Professor W. D. Verwey, *Humanitarian Intervention Under International Law*, Vol. XXXII NETH. INT'L L. REV. 357 (1985).

38. U.N. CHARTER art. 18.

39. U.N. CHARTER art. 28, ¶ 2.

40. U.N. CHARTER art. 11, ¶ 1.

adjustment of any situation . . . which it deems likely to impair the general welfare or friendly relations among nations."[41]

Under the "Uniting for Peace" resolution adopted by the General Assembly in 1950, the General Assembly may act if the Security Council—because of the exercise of a veto by one of the Permanent Five—fails to act in a case where there is a threat to the peace, a breach of the peace, or an act of aggression.[42] Then the General Assembly may make recommendations to members for collective action, including the use of armed force when necessary to maintain or restore international peace and security.[43] The General Assembly usually initiates studies, promotes "international cooperation in the economic, social, cultural, educational, and health fields, and assist[s] in the realization of human rights and fundamental freedoms for all,"[44] rather than authorizing coercive measures. The General Assembly receives reports through ECOSOC from the autonomous programs of the United Nations such as the Annual Report of the Governing Council of the United Nations Environment Program (UNEP).[45]

ECOSOC is the principal United Nations organ mandated to coordinate the economic and social work of the United Nations and to discuss international economic and social issues of a global or interdisciplinary nature.[46] ECOSOC's fifty-four members can formulate policy recommendations for member states and the United Nations system as a whole.[47] The specialized agencies are related to the United Nations system through individual agreements, as are nongovernmental organizations in consultative status.[48] A Standing Committee on Natural Resources attempts some coordination of environmental efforts through the different agencies. More work needs to be done in revitalizing ECOSOC.

Coordination for all issues within the United Nations system is handled through the Secretariat's Administrative Committee on Coordination (ACC), which consists of the heads of each specialized agency, the Secretary-General of the United Nations, and the heads

---

41. U.N. CHARTER art. 14.

42. G.A. Res. 377A, U.N. GAOR, 5th Sess., Supp. No. 20, at 10, U.N. Doc. A/1775 (1951).

43. *Id.* at 20.

44. U.N. CHARTER art. 13, ¶ 1 (b).

45. Catherine Tinker, Note, *The United Nations System: Prospects and Proposals for Reform, Environmental Planet Management by the United Nations: An Idea Whose Time Has Not Yet Come*, 22 J. INT'L L. & POL'Y 793, 822-23 (1990).

46. U.N. CHARTER art. 62.

47. *Id.*

48. U.N. CHARTER arts. 57, 63.

of the United Nations' quasi-autonomous agencies.[49] In the 1980s, a special committee called the Environmental Coordination Board met to address specific environmental activities, but it was disbanded after a few years because it duplicated the work of the ACC, and the Designated Officials on Environmental Matters (DOEM) replaced it.[50] Singling out environmental issues alone, especially when the system is trying to integrate environment and development, is counterproductive. Much more useful would be a strong voice on the ACC for environmental matters from a high-level executive head or director-general for sustainable development to make environment and development concerns part of every discussion of programs within the United Nations system.

IV. THE INTERNATIONAL COURT OF JUSTICE AND ARBITRAL TRIBUNALS

Arbitration to resolve conflicting values or competing uses for scarce resources, or adjudication to declare one course of action more equitable or appropriate than another is better than the use of force after harm has already occurred. These peaceful means of dispute resolution may be ordered by the Security Council without a finding of a threat to international peace and security[51] or be recommended by the General Assembly.[52] Dispute resolution can be pursued voluntarily by the parties involved or through diplomatic efforts to convince affected nations to resolve their disputes in this fashion.[53]

Until there is a clear body of international environmental law with definite rules of conduct for states, it is impossible to say that a state is in breach of its duties under international law. The United Nations Conference on Environment and Development negotiated a non-binding set of general principles of law on environment and development,[54] which may lead to the creation of customary inter-

---

49. Tinker, *supra* note 45, at 813 n.89.

50. Interview by author with John Washburn of the United Nations Secretariat, Executive Office of the Secretary-General, New York City, January, 1992.

51. U.N. CHARTER art. 36, ¶ 3.

52. U.N. CHARTER art. 14.

53. U.N. CHARTER art. 33, ¶ 1.

54. This document, initially entitled "Statement of General Rights and Obligations," or "Earth Charter," and finally entitled the "Rio de Janeiro Declaration on Environment and Development," was negotiated by member nations of the United Nations at a final preparatory meeting March 2-April 3, 1992, in New York at United Nations headquarters and adopted at the United Nations Conference on Environment and Development in Rio de Janeiro, Brazil, June 1-12, 1992.

For a full description of the contents of the penultimate draft of the document, U.N. Doc. A/Conf.151/PC/L.8/Rev.1, 31 August 1991, see Catherine Tinker, MAKING UNCED WORK: BUILDING THE LEGAL AND INSTITUTIONAL FRAMEWORK (1992). The final text is U.N. Doc. A/Conf.151/5/Rev. 1 (1992), *reprinted in* 31 INT'L LEGAL MATERIALS 874 (1992).

national law in time, or which may be included in new treaties. Until the principles included in the Rio Declaration become binding law, the only international environmental law universally accepted as customary international law is Principle 21 of the Stockholm Declaration, which states:

> States have, in accordance with the Charter of the United Nations and the principles of international law, the sovereign right to exploit their own resources pursuant to their own environmental policies, and the responsibility to ensure that activities within their jurisdiction or control do not cause damage to the environment of other States or of areas beyond the limits of national jurisdiction.[55]

Treaties, which state what duties parties are bound to observe and what constitutes a breach of those duties, usually include a dispute resolution mechanism. Many multilateral and bilateral environmental treaties call for arbitration and referral of disputes to the International Court of Justice (ICJ).[56] A few treaties create their own treaty bodies or conciliation commissions to resolve disputes arising under the treaty, such as the Canada-United States International Joint Commission[57] or the new Law of the Sea Tribunal.[58]

Under the United Nations Charter, the ICJ is the principal judicial organ of the United Nations.[59] The Security Council is directed by the Charter in Article 36(3) to "take into consideration that legal disputes should as a general rule be referred by the parties to the International Court of Justice."[60] The ICJ is competent under its statute to decide environmental disputes brought before it by states. The problem is the unwillingness of states to submit themselves to the jurisdiction of the court, rather than any weakness in the statute or organization of the court itself. Calls for reforming the ICJ or altering its statute to become an environmental tribunal are misplaced and unnecessary. One feature of the ICJ statute permits specialized chambers of the court to hear certain specialized types of cases,[61] making the ICJ quite attractive for environmental disputes. The flexibility of the chambers procedure, with five judges chosen by the parties compared to a hearing before the full court with fifteen judges, permits disputants to exercise political choice over the selection of the tribunal, allaying certain fears. The chambers procedure also allows the development of judicial expertise in the complicated

---

55. Declaration on the Human Environment, *supra* note 30.

56. *E.g.*, Convention on the International Trade in Endangered Species of Wild Fauna and Flora, Mar. 6, 1973, 27 U.S.T. 1087. (CITES).

57. The International Joint Commission was established in 1909.

58. Law of the Sea Convention, Dec. 20, 1982, Annex VI, *reprinted in* 21 INT'L LEGAL MATERIALS 1261 (1982) (U.N. Doc.A/Conf.62/122).

59. U.N. CHARTER art. 92.

60. U.N. CHARTER art. 36, ¶ 3.

61. I.C.J. STATUTE art. 26.

scientific and economic issues involved in environment and development disputes.

Short of adjudication by the ICJ, there are other avenues for settling disputes peacefully. Informal negotiations and diplomatic pursuits are the least coercive and, in any case, should be tried first, followed by conciliation or mediation efforts. Neutral third parties may be brought in at any point. Conciliation is a non-binding procedure where the third party is not actively involved; mediation, on the other hand, is characterized by direct involvement of the third party mediator, who seeks to bring the parties to a compromise. In addition, the Secretary-General of the United Nations may be asked to exercise the "good offices" function or to serve as a mediator or arbitrator.[62]

When any of these efforts fail, or the parties agree to do so, they may enter into arbitration. Here a single arbitrator or a panel of three arbitrators operating under one of several sets of international arbitration rules chosen by the parties hear arguments, examine evidence, and come to a decision which is usually binding.[63] The arbitrator acts more like a judge, finding for one side or the other and assessing damages or awarding other relief. In contrast, a mediator expresses personal opinions on the merits and tries to get both parties to give up something in order to reach an agreement. Both types of dispute resolution may be helpful in resolving resource disputes and other economic and social policies when conflicting values and goals are at stake in the international community.

V. CONCLUSION

Environmental hazards are threats to international peace and security. Adding environmental security to the agenda of the United Nations Security Council, however, is misplaced. To do so achieves little other than to create public awareness of the issue. Additionally, pushing environmental degradation or resource depletion onto the agenda of the United Nations Security Council creates the risk of escalating environmental disputes into larger military conflicts. Applying the definition of "threats to international peace and security, breaches of the peace and acts of aggression" to environment and development disputes may trigger Security Council action under Chapter VII (or even permit unilateral or collective self-defense) in situations where economic sanctions and military force may not be appropriate or effective.

---

62. FRANCK, *supra* note 22.

63. Commercial arbitration rules have been developed by various municipal or state law (such as London, Stockholm, and Switzerland) and by organizations such as the International Chambers of Commerce in Paris (ICC), the American Arbitration Association in New York (AAA), and other groups.

If the goal is to compel a nation to use alternatives to chlorofluorocarbons in manufacturing refrigerators or to discontinue overexploitation of scarce resources, finding economically or scientifically feasible methods that also protect the environment—including economic incentives that lead to voluntary compliance with international standards—is more likely to achieve that goal than the use of force. These problem-solving activities are outside the agenda of the United Nations Security Council and Chapter VII collective security actions. The General Assembly is the United Nations body currently responsible for issues affecting the environment and development and should continue that responsibility.

Mediation among parties with competing interests, arbitration, or adjudication by the ICJ are also far better alternatives than use of force by the Security Council. Future global conflicts will probably be caused by competing needs for dwindling supplies of natural resources, or by conflicting demands for short-term rewards from economic development using environmentally unsound practices versus the long-term benefits of preservation of air, water, or soil for present and future generations. Rather than call for Security Council consideration of "environmental security," the planet and the United Nations system would be better served by urging nations to develop strong international environmental laws and to take complaints to the ICJ when legal duties are breached or treaty obligations broken. Any incentives for voluntary compliance with techniques that promote sustainable development should be encouraged among policy-makers in governments, non-governmental organizations, industries, and businesses. Simultaneously, nations and organizations on the local and international level should work for the implementation of the recommendations of the United Nations Conference on Environment and Development and the further development of international environmental law.

Surface Water in the Iberian Peninsula: An Opportunity for Cooperation or a Source of Conflict?

JOSEPH W. DELLAPENNA*

> The challenge ahead for us is to transcend the self-interest of our respective nation-states . . . to embrace a broader self-interest—the survival of the human species in a threatened world.[1]

This article analyzes a particular situation involving transboundary waters and how the need to manage a scarce shared resource can become either a basis for cooperation or a festering source of ongoing and ever-worsening conflict. Water's unusual qualities of ambience and necessity have made it a frequent object of international controversy and conflict. Even the most cordially cooperative neighboring states have found it difficult to achieve mutually acceptable arrangements for their transboundary surface waters,[2] and states

---

* Professor of Law, Villanova University. LL.M. 1974, Columbia University; LL.M. in Public International and Comparative Law 1969, George Washington University; J.D. 1968, Detroit College of Law; B.B.A. 1965, University of Michigan. This article grew out of my work as a consultant to the Portuguese Directory-General of Natural Resources (Direcçao-Geral dos Recursos Naturais) as a Fulbright grantee in the summer of 1990. An earlier version of this paper was delivered at the World Jurists' Association Biannual Meeting in Barcelona on October 10, 1991. The analysis and conclusions are my own and do not necessarily represent the views of the American or Portuguese governments or of the Association.

1. Statement of Thomas McMillan, Canadian Minister of the Environment, before the World Commission on Environment & Development, Ottawa, Canada, May 26, 1986, *quoted in* Stephen McCaffrey, *International Organizations and the Holistic Approach to Water Problems*, 31 NAT. RESOURCES J. 139, 139 (1991).

2. Consider the apparently never ending disputes between the United States and Canada, notwithstanding the highly successful operations of the International Joint Commission on Boundary Waters. *See generally* JOHN KRUTILLA, THE COLUMBIA RIVER TREATY: THE ECONOMICS OF AN INTERNATIONAL RIVER BASIN DEVELOPMENT (1967); DON PIPER, THE INTERNATIONAL LAW OF THE GREAT LAKES (1967); Ralph Johnson, *The Columbia Basin*, in THE LAW OF INTERNATIONAL DRAINAGE BASINS 167 (Albert Garretson, *et al.* eds. 1967); Symposium, *U.S.-Canadian Transboundary Resource Issues*, 26 NAT. RESOURCES J. 201-376 (1986); Albert Utton, *Canadian International Waters*, in 5 WATERS AND WATER RIGHTS ch. 50 (Robert Beck ed. 1991). For studies of similar problems between other neighbors, see STEPHEN GOROVE, LAW AND POLITICS OF THE DANUBE (1964); NORRIS HUNDLEY, DIVIDING THE WATERS (1966); THE LAW OF INTERNATIONAL DRAINAGE BASINS, *supra*; JERRY MUELLER, THE RESTLESS RIVER (1975); LUDWIK TECLAFF, THE RIVER BASIN IN HISTORY AND LAW 152-184 (1967); U.N. DEV. AUTH'Y, RIVER BASIN DEVELOPMENT (1976); Albert Utton, *Mexican International Waters*, in 5 WATERS & WATER RIGHTS, *supra*, ch. 51.

within a single federal union have engaged in drawn out and bitter political and legal struggles over the waters they share.[3] Remarkably, with all of this struggle, the problems of transboundary aquifers have hardly begun to be considered.[4]

Conflicts over water have had a most unusual feature, at least in the twentieth century: No matter how violent other conflicts between states sharing a common watersource might have become, and especially when water itself played a central role in those conflicts, water facilities have remained off limits to combat, new

---

3. Consider, for example, the struggle between Arizona and California over the sharing of the lower Colorado River, a struggle made more difficult by the involvement of Mexico, other states of the United States, tribes of American Indians, and of other public and private interests. The dispute became a legal case under the original jurisdiction of the United States Supreme Court in 1929 after more than a decade of political struggle. Arizona v. California, 283 U.S. 423 (1931). The struggle, in both political and legal fora, continues today after no less than eight Supreme Court opinions. The most developed of these opinions is Arizona v. California, 373 U.S. 546 (1963). *See generally* Frank Trelease, *Arizona v. California: Allocation of Water Resources to People, States, and Nation*, 1963 SUP. CT. REV. 158. Colorado and Kansas have had an even longer-lasting struggle. Kansas v. Colorado, 475 U.S. 1079 (1986); Colorado v. Kansas, 320 U.S. 383 (1943); Kansas v. Colorado, 206 U.S. 46 (1907); Mark J. Wagner, Note, *The Parting of the Waters—The Dispute Between Colorado and Kansas Over the Arkansas River*, 24 WASHBURN L.J. 99 (1984). For similar, if less intense, battles between other states, see Texas v. New Mexico, 482 U.S. 124 (1987); Colorado v. New Mexico, 467 U.S. 310 (1984); Wisconsin v. Illinois, 388 U.S. 426 (1967); Texas v. New Mexico, 352 U.S. 991 (1957); New Jersey v. New York, 345 U.S. 369 (1953); Nebraska v. Wyoming, 325 U.S. 589 (1945); Wyoming v. Colorado, 309 U.S. 572 (1940); Washington v. Oregon, 297 U.S. 517 (1936); Nebraska v. Wyoming, 295 U.S. 40 (1935); New Jersey v. New York, 283 U.S. 336 (1931); Connecticut v. Massachusetts, 282 U.S. 660 (1931); Wisconsin v. Illinois, 281 U.S. 179 (1930); Wyoming v. Colorado, 259 U.S. 419 (1922); Missouri v. Illinois, 200 U.S. 496 (1906). *See generally* Douglas Grant, *Interstate Water Allocation*, in 4 WATERS AND WATER RIGHTS, *supra* note 2, at Pt. VIII; Bashir Hussain, *The Law of Interstate Rivers in India: Principles of Equitable Apportionment of River Waters*, 17 INDIAN J. INT'L L. 43 (1977); A. Dan Tarlock, *The Law of Equitable Apportionment Revisited, Updated, and Restated*, 56 COLO. L. REV. 381 (1985); Scott T. Anderson, Thomas A. Harder, Nancy Yost Laskaris, & Lori M. Mittag, Note, *Equitable Apportionment and the Supreme Court: What's So Equitable about Apportionment?*, 7 HAMLINE L. REV. 405 (1984).

4. Again, some of the most intense and illustrative examples come from the courts of the United States. *See, e.g.*, City of El Paso v. Reynolds, 597 F. Supp. 694 (D.N.M. 1984). *See generally* INTERNATIONAL GROUNDWATER LAW (Ludwik Teclaff & Albert Utton eds. 1981); Julio Barberis, *The Development of International Law of Transboundary Groundwater*, 31 NAT. RESOURCES J. 167 (1991); Robert D. Hayton & Albert E. Utton, *Transboundary Groundwaters: The Bellagio Draft Treaty*, 29 NAT. RESOURCES J. 663 (1989); International L. Ass'n, *International Rules on Groundwater*, REPORT OF THE SIXTY-SECOND CONFERENCE 21, 231-85 (Seoul, 1986); Ann Berkley Rodgers & Albert E. Utton, *The Ixtapa Draft Agreement Relating to the Use of Transboundary Groundwaters*, in TRANSBOUNDARY RESOURCES LAW 151 (Albert Utton & Ludwik Teclaff eds. 1987); Mary P. Keleher, Note, *Mexican-United States Shared Groundwater: Can It Be Managed?*, 1 GEO. INT'L ENVTL. L. REV. 113 (1988).

cooperative water arrangements have been negotiated, and preexisting arrangements have remained intact. For example, India and Pakistan engaged in three full-scale, albeit limited, wars since 1948, as well as numerous other skirmishes and serious threats of war. Yet in each instance they did not target water facilities or interfere in the operations of a joint Indo-Pakistani water management scheme specifically designed to survive the shock of war.[5]

Water has proven to be too critical a resource to fight over. Even when one state possessed a clear militarily dominance over its neighbors, that state has held back in taking all available water necessary for its needs through an apparent realization that depriving the populations of other states of essential water is one of the few steps that could make the weaker states desparate enough to fight against any odds and target the dominant state's own water facilities. Thus, even in the Jordan Valley in the midst of the Middle East's seemingly interminable low-level conflict and with occasional major wars over the last 50 years, tacit cooperation has been the almost unbroken rule between Israel and its neighbors, particularly Jordan.[6] Only in the recent Persian Gulf War did one side (the United States) target the water facilities of the other side (Iraq), in large part, no doubt, precisely because the militarily dominant state did not fear reciprocal attacks on its own domestic water facilities. The Draft Articles of the International Law Commission on the Law of Non-Navigational Uses of International Watercourses, recently transmitted to the United Nations General Assembly, would to some extent codify the practice of placing water sources and facilities off-limits to combat.[7]

Portugal and Spain have shared the water as well as the land of the Iberian peninsula for nearly nine centuries. After a long history of recurring conflict, they have avoided war between them since 1814. Today, both are members of the European Community, promising ever-closer ties across a broad range of activities, including the

---

5. *See* Brian E. Concannon, Note, *The Indus Waters Treaty: Three Decades of Success, Yet, Will It Endure?*, 2 GEO. INT'L ENVTL. L. REV. 55 (1989). *See also* TECLAFF, *supra* note 2, at 163-65, 183-84; Richard Baxter, *The Indus Basin*, in THE LAW OF INTERNATIONAL DRAINAGE BASINS, *supra* note 2, at 443.

6. *See generally* ADAM GARFINKLE, ISRAEL AND JORDAN IN THE SHADOW OF WAR 34-40, 79-83, 116, 162-73 (1992); Joseph W. Dellapenna, *Water in the Jordan Valley: The Potential and Limits of Law*, 5 PAL. Y.B. INT'L L. 15 (1989). *See also* ISRAEL AND ARAB WATERS: AN INTERNATIONAL SYMPOSIUM (Abdul Majid Farid & Hussein Sarriyeh eds. 1985).

7. *Draft Articles on The Law of the Non-Navigational Uses of International Watercourses*, Int'l L. Comm'n., U.N. GAOR, 43d Sess., at 7, U.N. Doc. A/CN.4/L.463/Add.4, ch. III, (1991), *reprinted in* 3 COLO. J. INT'L ENVTL. L. 1 (1992) [hereinafter Draft Articles]. Article 29 reads: "International watercourses and related installations, facilities and other works shall enjoy the protection accorded by the principles and rules of international law applicable in international and internal armed conflict and shall not be used in violation of those principles and rules."

gradual establishment of community-wide environmental policies.

The allocation of water within Portugal and Spain and even the maintenance of water-quality standards still remains largely a matter of national determination, subject only to traditional international law as a means of coordinating the countries' strongly conflicting policies. The European Community, thus far, has advised the two countries to resolve questions relating to shared water resources by themselves.[8] As will appear below, however, the present posture of national management of water leaves Portugal at a severe disadvantage *vis-à-vis* Spain and with only very limited means of persuading Spain to take Portuguese interests into account in Spanish management decisions.

Part I of this Article describes the political geography of Portugal and Spain, highlighting the problems presently facing them. Part II considers the treaties already implemented between Portugal and Spain and explains why those treaties do not meet the exigencies of the present situation. Part III examines the situation in the Iberian peninsula in light of international law. Part IV considers the practical strategies available to Portugal for persuading Spain to become more sensitive to Portuguese needs in planning and managing fresh water in Spain. Ultimately, I will argue the needs of the two nations can only be met by a system of joint management based on new institutions designed to represent and accommodate fairly the interests of both states.

I. THE POLITICAL GEOGRAPHY OF THE IBERIAN PENINSULA

The Iberian peninsula is divided between two nations, Portugal and Spain. Spain, Portugal's only neighboring state, is approximately five times the size of Portugal and has about four times as many people.[9] The Douro, Guadiana, and Tagus (Tejo) Rivers, three of the four major rivers of the peninsula, flow from Spain into Portugal; only the Ebro River is entirely within Spain.

In terms of national average precipitation, neither Spain nor Portugal is an especially water-poor country.[10] Precipitation, however, falls during a brief rainy season during the winter, with as much as 30 percent of the total sometimes falling during a single month.[11]

---

8. *See* text *infra* at notes 27-28, 34, 63-65, 118-28.

9. Spain's area is 504,750 square kilometers (194,884 square miles), with a population of nearly 40,000,000; Portugal's area is 91,985 sqare kilometers (35,516 sq. mi.), with a population of about 10,000,000. (The figures for both countries include several small off-shore island groups; the Spanish figures also include small areas in northern Africa.) THE STATEMAN'S Y.B. 1020, 1117 (Brian Hunter ed., 128th ed. 1991).

10. Joaquim Evaristo da Silva, Transboundary Water Resources Conflicts in the Iberian Peninsula 1 (1987) (unpublished monograph on file with the author).

11. *Id.*

The rain and snow are also concentrated in the more mountainous parts of the peninsula, leaving some low-land regions semi-arid, and in the case of southeastern Spain, a virtual desert. In Portugal, while precipitation averages about 900 millimeters (35 inches)/year, the range is from 3000 millimeters (120 inches)/year in some parts of the mountainous north to less than 500 millimeters (20 inches)/year in portions of the Algarve in the far south.[12]

The pattern of surface-water flow in the Iberian peninsula creates a measure of dependance on the Portuguese side which is not reciprocal: Approximately 70 percent of Portugal's surface supplies of fresh water come from rivers that arise in Spain. Discounting for flowage contributions from Portugal's share of the river basins, 40 percent or more of actual surface water in Portugal flows down from Spain.[13] Spain, which is currently using between 25 percent and 40 percent of its available water resources,[14] receives virtually none of its surface fresh water from Portugal and only negligible quantities from rivers flowing down from the French side of the Pyrenees.[15] Spain's only substantial risk relative to Portugal is from backflooding behind Portuguese dams. The treaties relating to water between the countries are agreements to consult before undertaking hydroelectric and other hydraulic projects that might affect similar projects in the other country.[16] This arrangement largely eliminates risks from unacceptable backflooding.

The topography of the peninsula, whereby Spain is virtually always upstream from Portugal, leaves Portugal vulnerable to injury from Spanish activities with no equivalent vulnerability on the part of Spain. Perhaps because the relationship is not reciprocal, an extensive history of cooperative management or even of negotiated sharing of the waters in question does not exist between the countries. Portugal currently faces shortages resulting both from its own rising demands for water and from rising water consumption in Spain. Intensifying water pollution coming from Spain and an apparent Spanish plan to place its only nuclear waste disposal site along the Duoro River just above the Spanish-Portuguese border add to Portugal's problems. On the other hand, Spanish works provide some benefits to water users in Portugal through regularizing the flows,

---

12. *Id.*

13. Maria de Fátima da Conceiçao Silva, Water Resources Planning and Irrigation in the Peninsula 1 (1983) (Working Paper, JNCIT/NAS/USAID Workshop on Water Resources at Ericeria, Portugal, April 13-15, 1983) (manuscript on file with author). Another author has estimated that 60 percent of Portugal's surface water flows down from Spain. Evaristo da Silva, *supra* note 10, at 1.

14. da Conceiçao Silva, *supra* note 13, at 4.

15. *See, e.g.*, The Lake Lanoux Arbitration (Spain v. Fr.), 24 I.L.R. 101 (1957), *digested in* 53 AM. J. INT'L L. 156 (1959).

16. See part II of this Article.

(potentially) reducing the flows in wet periods, and (more frequently) increasing flows during dry periods.[17] On balance, the problems appear to outweigh the benefits, especially as the operation of some Spanish dams, in fact, exacerbated flooding problems in Portugal rather than stemming floods through the regularization of flows.[18]

While Spain also must confront concerns about sharply rising demands for water resulting from by a growing population and rapid industrialization, Spain can do so largely without concern about activities in neighboring countries. Spanish uses, however, have reached a scale that threatens long-standing Portuguese uses and not simply the ability of Portuguese water users to initiate new uses.[19] Thus far, Spain has been unwilling to enter into discussions with Portugal about Spanish water management policies.

While irrigation works in the Iberian peninsula date back to antiquity, irrigation never existed to the extent now made possible by modern technology;[20] irrigation became, if anything, less widespread with the expulsion of the Moors.[21] Truly large-scale hydraulic works began only in the nineteenth century. As the peninsula's hydrologic data suggest would be necessary, large-scale hydraulic works focused on the storage of water for the dry season, initially for hydroelectric generation[22] and more recently for irrigation.[23] Later still, large-scale works were undertaken to transport water from storage sites in the wetter parts of the peninsula to the dryer parts. This activity took place within one country or the other, without significant cooperation between them apart from the sharing of information and consultations on the means of avoiding direct collisions between their works.[24]

Recent trends in urban and industrial development, occurring in both countries but more advanced in Spain, in the past two decades have made water managers in the peninsula more concerned about pollution and other environmental problems.[25] The managers now emphasize more efficient uses of water, particularly agricultural uses, rather than greater amounts for traditional patterns of use. Almost

---

17. da Conceiçao Silva, *supra* note 13, at 21-22.

18. See part I(B) of this Article. *See generally* Evaristo da Silva, *supra* note 10, at 6-7.

19. da Conceiçao Silva, *supra* note 13, at 3.

20. Ludwik A. Teclaff, *Fiat or Custom: The Checkered Development of International Water Law*, 31 NAT. RESOURCES J. 45, 63 (1991).

21. LUDWIK TECLAFF, WATER LAW IN HISTORICAL PERSPECTIVE 27 (1985); da Conceiçao Silva, *supra* note 13, at 6-7.

22. Evaristo da Silva, *supra* note 10, at 3.

23. da Conceiçao Silva, *supra* note 13, at 4, 6-18.

24. See part II of this article.

25. Evaristo da Silva, *supra* note 10, at 4. *See generally* George Matthew Silvers, Comment, *The Natural Environment in Spain: A Study of Environmental History, Legislation and Attitudes*, 5 TULANE ENVT'L L.J. 285 (1991).

certainly, integrated management of entire river basins will produce more efficient and more ecologically sound uses of the water than piecemeal development of isolated stretches of the rivers.[26] The joint institutional arrangements relating to the waters shared by the two nations are not designed to cope with such integrated approaches, and the two nations have not yet been able to work out, either directly or by way of their new memberships in the European Community, new arrangements suitable to their needs. The following three examples, one drawn from each of the major river basins shared between Portugal and Spain, illustrate the problems that have recently arisen relative to transboundary water management in the Iberian peninsula.

A. The Aldeadavilla Nuclear Waste Facility

The European Community requires Spain to indicate by 1999 one or more possible locations to store nuclear wastes from Spanish nuclear power plants. In September 1986, Spain proposed to the European Commission that Spain construct a nuclear waste laboratory on the Duoro River near the village of Aldeadavilla.[27] At Aldeadavilla, the Duoro forms the border between Spain and Portugal, shortly before entering Portugal to form the valley from which come Port wines, one of Portugal's major exports. The proposed facility will be less than one kilometer from the middle of the Duoro River, *i.e.*, from Portugal, and any contamination of the river will flow into Portugal. Even in the absence of actual contamination, the very presence of the project could socially and economically devalue the entire Portuguese share of the Duoro valley and its products. A group of experts from the Commission analyzed the project, and the Commission gave it financial support despite Portuguese objections.[28]

Ostensibly, the Aldeadavilla project is purely for research to test the behavior of the region's granite formations rather than to create a permanent storage site.[29] Granite is deeply fissured, and the Portuguese worry that even experimental work with nuclear wastes in the region will contaminate the Duoro River, particularly if the heat and pressure of drilling the burial chambers multiply or widen the fractures.[30] They ask why the Spanish have picked a research site so close to the Portuguese border when similar granite formations are found throughout Spain.[31] The Portuguese also fear that the research

26. da Conceiçao Silva, *supra* note 13, at 1.
27. Evaristo da Silva, *supra* note 10, at 4.
28. *Id.*
29. *Id.* at 5.
30. *Id.* at 6.
31. *Id.* at 5.

facility will become the permanent disposal site, pointing to the likely political pressures that will make a change in location difficult because the Aldeadavilla site threatens no community in Spain and building a nuclear-waste disposal facility elsewhere in Spain would threaten large Spanish communities.[32]

Portugal, a nonnuclear country, asks why the Portuguese must share any of the risk of disposing of another country's nuclear wastes.[33] While the European Parliament has voted that all nuclear waste facilities should be at least 100 kilometers from any international border,[34] the actions of the Parliament are not binding. Nor would making the resolution binding necessarily improve matters: Spain has the worst record of noncompliance with European Community environmental directives of any nation in the community.[35] Street demonstrations on both sides of the border to protest against the project also have had no effect.[36] Present institutional arrangements appear inadequate to provide an appropriate answer, and work on the Aldeadavilla project presses forward.

B. Flooding and Pollution of the Tagus (Tejo) River

The Tagus River basin includes about 30 percent of Portugal.[37] One of the major rivers of the Iberian peninsula, the Tagus is the site of numerous dams built by both countries for hydroelectric generation, flood control, agriculture, and public water supply. Because of the topography of the basin, the Spanish reservoirs are larger, with Portuguese reservoirs having relatively small storage capacities.[38] The Spanish reservoirs immediately upriver from Portugal are operated by private companies, which find it most profitable to keep their reservoirs as full as possible at all times. Thus Spanish dams seldom have any excess storage capacity available when flooding, a recurring and intensifying problem upstream from Lisbon, threatens.[39]

Unlike flooding in the Tagus valley, pollution problems within the Portuguese portion of the Tagus are largely a result of Portuguese discharges into the river.[40] Portuguese concerns about Spanish pollution of the Tagus, like the Aldeadavilla project on the Duoro, have focused on a nuclear facility. A Spanish nuclear power plant at

---

32. *Id.*
33. *Id.* at 6.
34. *Id.* at 5.
35. Silvers, *supra* note 25, at 285, 297, 306-09.
36. Evaristo da Silva, *supra* note 10, at 10.
37. *Id.*
38. *Id.* at 7.
39. *Id.*
40. *Id.*

Almaraz uses Tagus River water for cooling. Several years ago, a problem at the plant caused radioactivity all the way down to Lisbon.[41] Despite an agreement by Spain that its authorities would immediately notify Portugal of any problems with the nuclear plant, only one low-level official did so—unofficially—and he was, shortly thereafter, dismissed from his job.[42] While the problem did not become severe enough to suspend the water supply to Lisbon, the Portuguese are naturally apprehensive about the future of this plant.

C. The Alqueva Dam

The Portuguese are constructing the Alqueva Dam on the Guadiana River.[43] Designed to provide irrigation for between 135,000 and 200,000 hectares, hydroelectric generation, and urban and industrial water supply, the dam will be the largest in Portugal. It is the most controversial project in Portugal, having been the subject of intense debate stretching back over 20 years.[44] In addition to debate over the ecological effects of the project and its costs and benefits, the dam is exposed to Spanish diversion and pollution above the border. These concerns have exacerbated the controversy within Portugal over the Alqueva dam and seriously delayed its construction.[45]

The Guadiana River arises in Spain and, shortly after passing the city of Badajoz, briefly forms part of the international border between Spain and Portugal. As the river continues, Portugal becomes sovereign over both banks for a considerable stretch. The river again becomes an international border for its last reach before flowing into the Atlantic Ocean. The Alqueva Dam is to be located entirely within Portugal. Just upstream from where the river first serves as the border, Spain developed an irrigation project, supplying water to about 170,000 hectares and is considering an expansion of the irrigated area to 400,000 hectares.[46] The Spanish plans would seriously deplete the river's flow before it reaches the reservoir for the Alqueva Dam. Spain declared its intention to guarantee the minimum flow of the Guadiana during the dry periods, but the Spanish-Portugese

---

41. *Id.* at 8. Lisbon, Portugal's capital and largest city, is located near the mouth of the Tagus.

42. Evaristo da Silva, *supra* note 10, at 8.

43. da Conceiçao Silva, *supra* note 13, at 19; Evaristo da Silva, *supra* note 10, at 8.

44. da Conceiçao Silva, *supra* note 13, at 22; Evaristo da Silva, *supra* note 10, at 8-9.

45. Joaquim Evaristo da Silva, *Iberian International Rivers: Upstream Developments, Downstream Worries*, 5 TRANSBOUNDARY RESOURCES RPT. no. 2, at 4, 6 (Summer 1991).

46. Evaristo da Silva, *supra* note 10, at 9.

International Joint Commission, empowered by the 1968 Convention to do so, has been unable to establish the minimum flows or minimum annual volumes to which the guarantee is to apply.[47]

The International Joint Commission has failed to set minimum rates for the Guadiana because the Portuguese on the Commission have taken a very "soft" stance, exhibiting a typically Portuguese reluctance to confront Spain too directly.[48] In the meantime, Spanish farmers have begun to pump 10 m^3 of water directly from the border reaches of the river above the dam.[49] The Spanish are likely to claim their established uses as a legal right should the International Joint Commission ever get around to setting guaranteed minimums for the river.[50]

Spanish pollution of the Guadiana is related to the minimum flow problem. Badajoz dumps raw sewage into the Guadiana just above the point where it becomes the border.[51] Pollution is not yet covered by any convention between Spain and Portugal, and thus any guaranteed flow could well prove to be unusable in Portugal.

II. Applicable Treaties

Spain has always been very cautious about entering into international agreements or arrangements that might compromise Spanish sovereignty over its resources.[52] Given Spain's consistent upstream situation, it has seldom been in a position to benefit from acknowledging any downstream rights. Portugal, on the other hand, consistently has been reluctant to challenge Spain on water issues, in part because this is a common pattern in the Portuguese approach to their much larger neighbor,[53] but also because Portugal's uniformly downstream situation offers few obvious bases from which to bargain over water. As a result, the Portuguese have tended only to seek information about Spanish developments while concentrating on building hydroelectric and other hydraulic works in order to better exploit the water available within Portugal.[54]

Despite the foregoing reasons militating against agreements on shared waters, Portugal and Spain have entered into four agreements affecting water issues. The first was a convention between Portugal

---

47. *Id.* The 1968 Convention allocated 4 BCM/yr. (billion cubic meters/year: 4,000,000,000 m^3/yr.). *See infra* text accompanying note 62.
48. Evaristo da Silva, *supra* note 10, at 9.
49. *Id.*
50. *Id.* at 10.
51. *Id.*
52. *Id.* at 3.
53. *Id.* at 9.
54. *Id.*

and Spain signed in 1866.[55] This convention requires consultations before either signatory would license a private hydraulic work on the international reaches of transboundary rivers.

Second, a convention signed in 1927 divided the international portion of the Duoro River into two parts, allowing Spain to exploit the hydroelectric potential of the first part and Portugal the hydroelectric potential of the second part.[56] The 1927 convention also contains guarantees of minimum flows[57] and establishes an International Joint Commission to share information about the development of the hydroelectric potential of the international reaches of the transboundary rivers.[58] The Commission is empowered to decide whether proposed works are compatible with the convention's provisions; unanimous decisions are immediately binding on the parties, but majority decisions must be approved by the two governments, with approval presumed if neither government objects within 30 days of the communication of the decision to the governments.[59] The convention also provides, theoretically, for recourse to the International Court of Justice should the parties fail to agree;[60] the agreement, however, makes no provision for the implementation of a judicial award.

Third, an agreement in 1964 extended the authority of the International Joint Commission over other sorts of hydraulic works and introduced a measure of flexibility in the sharing of the hydroelectric potential of the Duoro River.[61] In light of the Portuguese project at Alqueva, the powers of the International Joint Commission to guarantee minimum flows were extended to the Guadiana River in 1968.[62]

Finally, Spain and Portugal joined the European Community on January 1, 1986.[63] Both countires are phasing in community standards relative to the quality of their waters, reinforcing the existing tendency, noted above, to make more efficient and less damaging uses of water.[64] Thus far the Community institutions appear neither to have had any impact on the allocation of water within or between

---

55. Agreement on Regulations of Boundary Waters, November 20, 1866, as an Annex to the Convention on Boundaries, September 29, 1864, Spain-Port., 129 Consol. T.S. 453.

56. Convention to Regulate the Hydro-Electric Development of the International Section of the River Duoro, Aug. 11, 1927, Spain-Port., 82 L.N.T.S. 133, art. 2.

57. *Id.*, arts. 8, 18.

58. *Id.*, art. 14.

59. *Id.*, art. 16.

60. *Id.*, art. 21.

61. da Conceiçao Silva, *supra* note 13, at 19.

62. *Id.* See part II(C) of this Article.

63. Evaristo da Silva, *supra* note 10, at 4.

64. *See supra* text accompanying note 18.

the two nations nor to have offered any protection to Portugal from degradation of shared waters by activities in Spain.[65]

III. THE CUSTOMARY INTERNATIONAL LAW OF TRANSBOUNDARY RIVERS

In the absence of express international agreements, international law operates through customary international law, which consists of the practices that states engage in out of a sense of legal obligation (the *opinio juris*).[66] Customary law (regional or general) develops through a process of claim and counterclaim between states;[67] practices that crystallize as customary international law can include treaties or other agreed arrangements,[68] informal decisions reflected by votes in international assemblies,[69] decisions by courts or international arbitrators,[70] or unilateral actions. The writings of well-respected scholars of international law (termed "the most highly qualified publicists" in the Statute of the International Court of Justice)[71] often contain the best evidence of what those practices are and whether those practices arise from an *opinio juris* or from other motives unrelated to law.

Customary international law, in its current state of somewhat primitive development,[72] cannot solve the management problems con-

---

65. Evaristo da Silva, *supra* note 10, at 4. The European Community has indicated that Spain has the worst record for noncompliance with community environmental directives of any member state. Silver, *supra* note 25, at 285, 297, 306-09.

66. MYRES MCDOUGAL, HAROLD LASSWELL, & IVAN VLASIC, LAW AND PUBLIC ORDER IN SPACE 116 (1963).

67. WATER IN THE MIDDLE EAST 158-62, 167 (Thomas Naff & Ruth Matson eds. 1984). The classic description of this process is found in Myres M. McDougal & Norbert A. Schlei, *The Hydrogen Bomb Test in Perspective: Lawful Measures for Security*, 64 YALE L.J. 648 (1955). *See also* CHARLES DE VISSHER, THEORY AND REALITY IN INTERNATIONAL LAW (1968).

68. Treaties to which a particular state is not a party might be evidence of a custom binding on that state. *See* MCDOUGAL, LASSWELL, & VLASIC, *supra* note 66, at 82-82, 115-19; JULIUS STONE, LEGAL CONTROLS IN INTERNATIONAL LAW 135 (1954). *But see* FRIEDRICH BERBER, RIVERS IN INTERNATIONAL LAW 128-37 (R.K. Bastone trans. 1959); 1 CHARLES HYDE, INTERNATIONAL LAW 12 (2d ed. 1945).

69. *See* Christopher Joyner, *U.N. General Assembly Resolutions and International Law: Rethinking the Contemporary Dynamics of Norm-Creation*, 11 CAL. W. INT'L L.J. 445, 477 (1981).

70. *See generally*, HERSCH LAUTERPACHT, THE DEVELOPMENT OF INTERNATIONAL LAW BY THE INTERNATIONAL COURT (1958); 2 SHABTAI ROSENNE, THE LAW AND PRACTICE OF THE INTERNATIONAL COURT 611-13 (1965); Michael Akehurst, *The Hierarchy of Sources in International Law*, 47 BRIT. Y.B. INT'L L. 273 (1975).

71. Statute of the International Court of Justice, 59 Stat. 1055, T.S. 993, art. 38(1)(d) (1945).

72. H.L.A. HART, THE CONCEPT OF LAW 77-96 (1961); WATER IN THE MIDDLE EAST, *supra* note 67, at 157-160; Yoram Dinstein, *International Law as a Primitive Legal System*, 19 INT'L L. & POLITICS 1 (1986).

fronting Spain and Portugal. Yet such customary law is not wholly without utility: Customary international law both empowers international actors by legitimating their claims and limits them by circumscribing the kinds of claims they are permitted to make. In the absence of an enforcement mechanism, however, international law has nothing better to offer than the law of the vendetta.[73]

A. The Customary International Law of Transboundary Rivers Generally

Space permits only a summary description of the customary international law applicable to shared surface-water bodies.[74] For nonnavigational uses of water, international claims and counterclaims have followed a predictable pattern, depending on the riparian status of the state making the claim. To begin with, all states agree that only riparian states—states across which, or through which, a river flows—have any legal right, absent agreement, to use the water of a river.[75] Beyond that simple point, however, the patterns of claim and counterclaim initially diverge sharply according to whether the claimant state is an upper or a lower riparian.

The uppermost-riparian states initially base their claims on "absolute territorial sovereignty,"[76] typically claiming the right to do whatever they choose with the water regardless of its effect on other riparian states. Downstream states, on the other hand, begin with a claim to the "absolute integrity of the river,"[77] claiming that upper-riparian states can do nothing that affects the quantity or quality of water that flows down to them. The utter incompatability of such claims guarantees that neither claim will prevail, but the process

---

73. WATER IN THE MIDDLE EAST *supra note 67 at 161. See also* Richard B. Bilder, *Some Limitations of Adjudication as an International Dispute Settlement Technique*, 23 VA. J. INT'L L. 1, 4-5 (1982); Richard A. Falk, *The Beirut Raid and the International Law of Retaliation*, 63 AM. J. INT'L L. 415, 442 (1969).

74. For illustrative works on the law of transboundary surface waters, see Draft Articles *supra* note 7; BERBER, *supra* note 67; BRIJ CHAUHAN, SETTLEMENT OF WATER LAW DISPUTES IN INTERNATIONAL DRAINAGE BASINS (1981); GEORGES KAECKENBEECK, INTERNATIONAL RIVERS (1962); Richard B. Bilder, *International Law and Natural Resources Policies*, 20 NAT. RESOURCES J. 451 (1980); Jan Hostie, *Problems of International Concerning Irrigation of Arid Lands*, 31 INT'L AFFAIRS 61 (1955); Teclaff, *supra* note 20; Albert Utton, *International Streams and Lakes Generally*, *in* 5 WATERS AND WATER RIGHTS, *supra* note 2, ch. 49.

75. WATER IN THE MIDDLE EAST, *supra* note 67, at 166-67.

76. *Id.* at 164-65. This theory has one of its best known expressions in a published opinion by United States Attorney General Harmon, 21 Op. Att'y Gen. 274, 281-82 (1898). The "Harmon Doctrine" has since been disapproved by the United States. State Department, *Memorandum to the Legal Advisor*, Nov. 23, 1942, in 3 MARJORIE WHITEMAN, DIGEST OF INTERNATIONAL LAW 950-54 (1964).

77. WATER IN THE MIDDLE EAST, *supra* note 67, at 165; A. P. Lester, *River Pollution in International Law*, 57 AM. J. INT'L L. 828, 832 (1963).

might take decades of negotiations or worse to work out a solution.

The solution initially is found in a concept of "restricted sovereignty."[78] Lower riparian states, particularly those wedged along a river so as to be both upper and lower riparians on the same stream, often adopt the theory of restricted sovereign rights under which each state recognizes the right of all riparian states to use some water from a common source and the obligation to manage their uses so as not to interfere with similar uses in other riparian states. The quantity of water to which each state is entitled might be defined according to some historic pattern of use, although occasionally some other more-or-less objective measure of need is advanced (such as population, area, or arable land). On the other hand, the definition might be no more developed than the vague notion that each state is entitled to a "reasonable share" of the water.

Eventually some *modus vivendi* is worked out on most international river systems based on the notion of restricted sovereignty—nearly 100 such treaties had entered into force by 1950, and more have followed.[79] International judicial and arbitral awards are to a like effect.[80] The respected publicists of international law are in virtual unanimous agreement on the same point.[81]

Some international agreements relating to shared water resources embrace a concept that might be described as a "community of

---

78. WATER IN THE MIDDLE EAST, *supra* note 67, at 165-66.

79. *Report of the U.N. Commission for Europe, Legal Aspects of Hydro-Electric Development of Rivers and Lakes of Common Interest*, at 95-152, U.N. Doc. E/ECE/136 (1952); BERBER, *supra* note 67; Herbert Smith, THE ECONOMIC USES OF INTERNATIONAL RIVERS (1931); Utton, *supra* note 74, at § 49.03(a).

80. *See, e.g.*, Case of the Territorial Jurisdiction of the Int'l Comm'n of the Oder River, [1929] P.C.I.J., ser. A, No. 23 at 27; The Lake Lanoux Arbitration (Fr. v. Spain), 24 I.L.R. 101, 139 (1957), *digested in* 53 AM. J. INT'L L. 156, 170 (1959). *See generally* Utton, *supra* note 74, at § 49.03(b).

81. *See generally* International L. Assoc., *The Helsinki Rules on the Uses of the Waters of International Rivers* (REP. OF THE 52D CONF., ADOPTED AT HELSINKI, AUG. 20, 1966) [HEREINAFTER HELSINKI RULES]; BERBER, *supra* note 68, at 25, 272-74; DANIEL O'CONNELL, INTERNATIONAL LAW 556-58 (2d ed. 1970); 1 LASSA OPPENHEIM, INTERNATIONAL LAW 474-75 (8th ed., Hersch Lauterpacht ed. 1955); SMITH, *supra* note 79, at 150-51; TECLAFF, *supra* note 2, at 152; Dominique Alheritiere, *Settlement of Public International Disputes on Shared Resources: Elements of a Comparative Study of International Instruments*, in TRANSBOUNDARY RESOURCES LAW, *supra* note 4, at 139-49; Juraj Andrassy, *L'Utilisation des Eaux des Bassins Fluviaux Internationaux*, 16 REVUE EGYPTIENNE DE DROIT INTERNATIONAL 23 (1960); Dante Caponera, *Patterns of Cooperation in International Water Law*, in TRANSBOUNDARY RESOURCES LAW, *supra* note 4, at 1, 3-10; Aziza Fahmi, *International River Law for Non-Navigable Rivers with Special Reference to the Nile*, 23 REVUE EGYPTIENNE DE DROIT INTERNATIONAL 39 (1967); Gretta Goldenman, *Adapting to Climate Change: A Study of International Rivers and Their Legal Arrangements*, 17 ECOL. L.Q. 741, 741 (1990); Sayed Hosni, *The Nile Regime*, 17 REVUE EGYPTIENNE DE DROIT INTERNATIONAL 70, 70 (1961); Utton, *supra* note 74, § 49.03(e).

property" in the watersource.[82] Under the community of property concept, the waterbasin is jointly developed and managed as a unit without regard to international borders and with an agreed sharing of the benefits of, and equitable participation in, that development and management.[83] Although the full instantiation of such an approach is still rare,[84] good reasons exist for believing that the practice of nations will move in this direction.

Even if each actor were to agree to the concept of water as a shared resource requiring recognition that the sovereignty of each riparian state is limited relative to the water, states would still dispute what should be the common standard and the proper application of the agreed standard. Such disputes could ultimately lead back to the law of the vendetta. Serious conflict in one form or another cannot be avoided without a peaceful mechanism for orderly investigating and resolving the inevitable disputes that are the distinguishing characteristic of the restricted-sovereignty theory. These conflicts will undoubtedly push nations towards the model of a community of property approach to shared water resources.[85]

The concept of an international drainage basin is widely supported by naturalists, engineers, and economists, as well as jurists.[86] Ludwik Teclaff elaborated the concept in his well-known book entitled *The River Basin in Law and History*.[87] In recent years, a number of international meetings have adopted the principle of community of property or equitable participation in shared water resources, culminating in the recently completed Draft Articles of the International Law Commission on the Law of the Non-Navigational Uses of International Watercourses.[88]

B. The Pronouncements of Quasi-Public International Organizations

Virtually every international organization to have considered the matter has endorsed either restricted sovereignty or community of property as a rule of customary international law. One of the best

---

82. Utton, *supra* note 74, at § 49.03.

83. L.F.E. Goldie, *Equity and the International Management of Transboundary Resources*, in TRANSBOUNDARY RESOURCES LAW, *supra* note 4, at 103-37.

84. See, *e.g.*, sources collected at note 2 on Canadian-United States boundary waters. *See also* The Treaty for Amazonian Co-operation, art. I, *reproduced in* 17 INT'L LEG. MATERIALS 1045, 1046 (1978); Jens Evenson, *Third Report on the Law of Non-Navigational Watercourses*, [1982] II Y.B. Int'l L. Comm'n 80-81, U.N. Doc. A/CN.4/348.

85. WATER IN THE MIDDLE EAST, *supra* note 67, at 171-73.

86. McCaffrey, *supra* note 1, at 143.

87. TECLAFF, *supra* note 2.

88. Draft Articles, *supra* note 7, art. 8. *See generally* Utton, *supra* note 74, at § 49.09.

known of such actions was by the International Law Association, a nongovernmental organization of legal experts founded in 1873.[89] In 1954, the Association undertook a project to codify the law relating to the shared uses of international rivers. The result was the "Helsinki Rules on the Uses of the Waters of International Rivers," adopted in 1966.[90] The Helsinki Rules were the first attempt by any international organization to codify the entire law of international watercourses.[91]

The Helsinki Rules center on the concept of international drainage basins (watersheds extending over two or more states) as an indivisible hydrologic unit on the basis of which planning must occur to assure the "maximum utilization and development of any portion of its waters."[92] The Helsinki Rules explicitly include within this concept all tributaries (including tributary groundwater) and not simply the international watercourse itself.[93] Within a drainage basin, the Helsinki Rules embrace the concept of restricted sovereignty through adoption of a rule of "equitable utilization."[94] The International Law Association has continued to draft rules relating to water-centered activities not addressed directly by the Helsinki rules, including rules relating to flood control (1972), pollution (1972 & 1982), navigability (1974), the protection of water installations during armed conflicts (1976), joint administration (1976 & 1986), flowage regulation (1980), general environmental management concerns (1980), and groundwater (1986).[95]

Other public and quasi-public international organizations have made similar pronouncements, including the Institut de Droit International, the Inter-American Bar Association, and the New York University Research on International Law.[96] The ongoing work of the International Law Association developed what some claim to be a fifth principle governing the management of shared water resources,[97] that each nation ensure that acts within that nation not

89. McCaffrey, *supra* note 1, at 141.

90. Helsinki Rules, *supra* note 81.

91. McCaffrey, *supra* note 1, at 141.

92. Helsinki Rules, *supra* note 81, at 7-8 (art. II & comment (a)).

93. *Id.* at 7-8.

94. *Id.*, art. IV. The phrase "equitable utilization" is similar in both phrasing and in meaning to the rule of "equitable apportionment" applied by the Supreme Court of the United States to interstate disputes over surface waters shared between the disputing states—a system that has barely functioned in a society with a strong judicial structure to resolve disputes between users. *See* sources collected at note 3.

95. *See generally* McCaffrey, *supra* note 1, at 144-50.

96. Institut de Droit International, *Utilization of Non-Maritime International Waters (Except for Navigation)*, ART. 2 (SEPT. 4-13, 1961); INTER-AMERICAN BAR ASS'N, *Resolution on Principles of Law Governing the Uses of International Rivers and Lakes* (1957); International Law Ass'n, *Principles of Law and Recommendations on the Uses of International Rivers* 197-98 (1958) [hereafter N.Y.U. Conference].

97. For the first four approaches, see part III(A) of this Article.

cause "substantial damage" to the environment or the natural condition of the waters beyond the limits of the nation's jurisdiction.[98] Section 601 of the *Restatement (Third) of Foreign Relations Law* also declares that states must "take such measures as may be necessary, to the extent practicable under the circumstances" to avoid injury to neighboring states.[99]

The International Law Commission, an organ of the United Nations, in its Draft Articles submitted to the General Assembly in 1991 embraced both the principle of equitable apportionment and the obligation not to cause appreciable harm to other states. The relevant draft rules read as follows:

> *Article 5*
> *Equitable and reasonable utilization and participation*
> (1) Watercourse States shall in their respective territories utilize an international watercourse in an equitable and reasonable manner. In particular, an international watercourse shall be used and developed by watercourse States with a view to attaining optimal utilization thereof and benefits therefrom consistent with adequate protection in the watercourse.
> (2) Watercourse States shall participate in the use, development and protection of an international watercourse in an equitable and reasonable manner. Such participation includes both the right to utilize the watercourse and the duty to cooperate in the protection and development thereof, as provided in the present articles.
> . . . .
> *Article 7*
> *Obligation not to cause appreciable harm*
> Watercourse States shall utilize an international watercourse in such a way as not to cause appreciable harm to other watercourse States.[100]

Even though the Chief Rapporteur for the project in 1982 acknowledged the virtually unanimous recognition of the rule of "equitable utilization" as a general rule of international law,[101] Stephen McCaffrey, the final Rapporteur for the project, concluded that the International Law Commission intended the rule of no appreciable harm to be superior to the rule of equitable sharing.[102] Perhaps one

---

98. *See, e.g.*, International L. Ass'n, *Rules on the Relationship between Water, Other Natural Resources and the Environment*, art. I (adopted at Belgrade, 1980).

99. RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW § 601 (1987). *See also* N.Y.U. Conference, *supra* note 96, at 197.

100. Draft Articles, *supra* note 7, arts. 5, 7.

101. Evenson, *supra* note 84, at 85. *See also* McCaffrey, *supra* note 1, at 150-61.

102. Stephen C. McCaffrey, *The Law of International Watercourses: Some Recent Developments and Unanswered Questions*, 17 DEN. J. INT'L L. & POL'Y 505, 509-10 (1989).

can reach this conclusion based on comparing the categorical command in article 7 to the more precatory language of article 5, but this conclusion ignores the express provisions of the Draft Articles:

> *Article 10*
> *Relationship between uses*
> 1. In the absence of agreement or custom to the contrary, no use of an international water course enjoys priority over other uses.
> 2. In the event of a conflict between uses of an international water course, it shall be resolved with reference to the principles and factors set out in articles 5 to 7, with special regard being given to the requirements of vital human needs.[103]

To assert the absolute primacy of the rule of no appreciable harm also ignores the reality of water usage. Carried to its logical extreme, a principle of no appreciable harm would prohibit any meaningful use by an upper-riparian state, turning the no-harm rule into merely a variant form of the absolute- integrity claim. That position, while frequently advocated by lower-riparian states, has never in fact been adopted by actual international decision-makers.[104]

One can reconcile the two rules by stressing that the no- harm rule has been variously stated as prohibiting "appreciable harm," "sensible harm," "significant harm," "substantial harm," or the like.[105] Whether harm exceeds one or another of these standards can be determined only by considering whether a use represents a reasonable or equitable utilization,[106] for as the German federal supreme court stated in the *Danauversinkung Case (Württemberg v. Baden)*,[107] "[o]ne must consider not only the absolute injury caused to the neighboring State, but also the relation of the advantage gained by one to the injury caused to the other."[108] By this view, the rule of no appreciable harm is really just a variant statement of the rule of equitable apportionment or equitable participation under the principle of restricted sovereignty or the principle of community of property in the watersource.[109]

---

103. Draft Articles, *supra* note 7, art. 10. Article 6 describes, in highly general terms, the factors to be considered in determining whether a use is reasonable and an apportionment is equitable.

104. *See supra* text accompanying note 77.

105. Evenson, *supra* note 84, at 98-100.

106. *Id.*, at 100-10; Helsinki Rules, *supra* note 81, at 19-20 [commentary to Art. X]; International L. Ass'n, *supra* note 98, art. 1. *See generally* McCaffrey, *supra* note 1, at 144-50; Utton, *supra* note 74, at §§ 49.04, 49.10.

107. Ann. Digest & Rep. of Pub. Int'l L. Cases 128 (RGst. 1927). *See also* Evenson, *supra* note 84, at 102.

108. *See generally* Utton, *supra* note 74, at §§ 49.05, 49.06

109. *See generally* R. H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960).

C. The Spanish Precedent

The leading Portuguese text on the international water management unequivocally endorses the principle of equitable sharing of transboundary waters.[110] Conversely, Spain, *vis-à-vis* Portugal, has embraced the typical upper-riparian state's claim of absolute territorial sovereignty.[111] Today, however, there can be little doubt that the present Spanish approach utterly ignores Portugal's right to an equitable share of the water under general customary international law, and as such constitutes a legal wrong. The Spanish claim to absolute sovereignty over waters within Spain is even more vulnerable than such claims by other upper-riparian states because Spain itself has successfully espoused the rule of the absolute integrity of the river in *The Lake Lanoux Arbitration (Spain v. France)*.[112]

Lake Lanoux is a small lake, located entirely in France; from Lake Lanoux, a small river flows into the Carol River, which flows into Spain. The French government proposed to divert the waters of the Carol River over a precipitous 780-meter drop into the Ariege River to generate electricity. Originally, France claimed the right of absolute sovereignty as its basis for doing so. When Spain complained that this project could not be undertaken without its consent, the French eventually promised to divert water (equivalent in volume and quality) downstream from the project from the Ariege River to replenish the Carol River before it entered Spain. France and Spain agreed to arbitration to determine whether the proposed action would violate Spanish rights.

Because of the plan to restore the Carol River as to both the quantity and the quality of its waters before the river entered Spain, the arbitration panel held that the planned works would not violate either customary international law or Spanish rights under the Treaty of Bayonne[113] by which the two nations had agreed to coordinate hydroelectric development of their shared waters. In reaching this conclusion, the tribunal expressly declared that the rule of international law relative to shared water resources was the rule here termed restricted sovereignty.[114] While a common lawyer might be inclined to view this pronouncement as mere dictum, international tribunals operate in the civil law tradition, which while denying a binding

---

110. LUIS VEIGA DA CUNHA, VITO ALVES DE FIGUEIREDO, MÁRIO LINO CORREIA, & ANTÓNIO DOS SANTOS GONÇALVES, MANAGEMENT AND LAW FOR WATER RESOURCES 211-24, 241-43 (1977).

111. *See supra* text accompanying notes 76-77.

112. The Lake Lanoux Arbitration (Fr. v. Spain), 24 I.L.R. 101 (1957), *digested in* 53 AM. J. INT'L L. 156 (1959).

113. Signed, May 29, 1866, 56 Brit. & For. State Papers 212.

114. Lake Lanoux Arbitration, *supra* note 112, 101 I.L.R. at 139, 53 AM. J. INT'L L. at 170.

force to judicial or arbitral decisions that follow from a strict rule of precedent, accords all parts of the opinion equal weight as the teaching of "most highly qualified publicists."[115]

D. An Aside on Groundwater

The Helsinki Rules include only those groundwaters that formed part of a drainage basin, that is, that contributed to the principle streams, lakes, or other common terminus of the relevant watershed.[116] While there is far less experience regarding disputes over aquifer management, the same principles would no doubt be applied by analogy.[117] A gathering of experts on the law of international water recently confirmed this conclusion in a meeting at Bellagio, Italy, where they drafted a model treaty to assure the equitable utilization and management of shared groundwater basins.[118]

IV. Dispute-Settlement Mechanisms

Under present circumstances, intense disputes appear to be inevitable. Conflict would be likely even if the existing agreements between Portugal and Spain tied the consumption of water more effectively to an objective measure of need (such as historic use or arable acreage). The situation is even worse, however, as the rights of the two states have been largely left to measurement by the vaguely defined standard of equitable utilization tempered by the right not to be appreciably harmed. While Spain seems to be in clear violation of these norms, Portugal has limited means for redressing these wrongs.

Consider an apparently simple matter such as the inventorying of the water available to a country. Effective management of the water by a nation is impossible without a correct inventory of the quantity and quality of available water. For Portugal, such an inventory is impossible without substantial Spanish cooperation.[119] Spanish water managers, as is usual where Portugal is concerned, do not require foreign cooperation to inventory Spanish waters. Furthermore, having adopted a modern water law in 1985,[120] Spain's

---

115. *See supra* text accompanying note 71.
116. Helsinki Rules, *supra* note 81, at 8 (comment (b)).
117. *See* sources collected at note 4.
118. Hayton & Utton, *supra* note 4.
119. Evaristo da Silva, *supra* note 10, at 4.
120. Ley 29/1985, de 2 de agosto, de aguas. *See* Ministro de Obras Públicas y Urbanismo, *Proyecto de Ley de Aguas* (1985) (copy on file with the author). *See also* Silvers, *supra* note 35, at 302-03.

system of water management is both legally and materially more advanced than Portugal's.[121]

Neither existing agreements between Spain and Portugal nor customary international law will yield definite rules about what either nation can or cannot do relative to their shared rivers. Nor have the institutions of the European Community been willing to turn their attention to the problems of water management in the Iberian peninsula. The European Community first began to take an interest in environmental affairs in the early 1970s as a means of preventing environmental regulations from serving as hidden trade barriers, but subsequently has come to be concerned about protecting and improving the natural environment.[122] By now, the Community has issued some 150 directives, regulations, and decisions relating to the environment, but nearly all have concerned the setting of uniform minimum standards for environmental quality. The Community organs have virtually nothing to say about allocating scarce resources across national boundaries.[123] The recent creation of a European Environmental Agency is unlikely to change this pattern as that agency is merely empowered to gather data and to establish standards for the reporting of data.[124]

The Single European Act amended the Treaty of Rome in 1987 to provide that ensuring a prudent and rational utilization of natural resources was one of the legislative competencies of the European Community.[125] This provision resolves any questions about the competence of the Community to involve itself in resource allocation decisions, and such decisions, when made, will clearly prevail over inconsistent national laws.[126] Still, the clear focus of the new treaty provisions remains on the prevention of environmental damage rather than the allocation of shared resources.[127]

---

121. Portugal's water laws date back to 1919, although a project is currently underway to revise them. Decreto no. 5787-IIII, de 10 de Maio de 1919 (copy on file with the author).

122. Thomas Bunge, *European Environmental Law: Community Legislation and Member States' Competences Under the EEC Treaty*, 59 REV. JUR. U.P.R. 669, 670 (1990). *See generally* ENVIRONMENTAL LAW OF THE EUROPEAN COMMUNITIES (W. Burhenne ed. 1990); E. REHBINDER & RICHARD STEWART, ENVIRONMENTAL PROTECTION POLICY (1985); Christian Zacker, *Environmental Law of the European Economic Community: New Powers Under the Single European Act*, 14 B.C. INT'L & COMP. L. REV. 249, 261-64 (1991); Gerard V. Curtin, Jr., Note, *Regulation 1210/90: Establishment of the European Environmental Agency*, 14 B.C. INT'L & COMP. L. REV. 321, 321-25 (1991).

123. Bunge, *supra* note 122, at 670-71.

124. Regulation 1210/90, Council Regulation of May 7, 1990, O.J. L120/1 (1990). *See generally* Curtin, *supra* note 122, at 325-31.

125. 30 O.J. L169/1, arts. 130r, 130s (1987) [hereafter Single European Act].

126. Bunge, *supra* note 122, at 676-79, 683-90. Remember, however, that Spain has the worst record of noncompliance with European Community environmental directives. Silvers, *supra* note 35, at 285, 297, 306-09.

127. Single European Act, *supra* note 125, art. 130r(2).

Should the European Community turn its attention to resource allocation disputes, the lengthy process necessary to adopt any new directive, regulation, or decision assures that no prompt results can be expected from this quarter.[128] Furthermore, decisions relating to environmental management must be unanimous, although the Council of the Community is empowered to define—by unanimous vote—issues which can be decided by a qualified majority rather than unanimity.[129] This unanimity requirement leaves allocation resolutions firmly in the hands of the contending states.

The Economic Commission for Europe, operating under the auspices of the United Nations, also does not provide a water-allocation or a dispute-resolution mechanism. The Commission has adopted three instruments relative to international water management. The "Declaration of Policy on Prevention and Control of Water Pollution, Including Transboundary Pollution," merely indicates that "rational utilization of water resources" is to be basic element of long-term water management.[130] This declaration was followed by a "Declaration of Policy on the Rational Use of Water," which recommended a "unified strategy" and "coordinated utilization."[131] Finally, the Commission adopted "Recommendations to ECE Governments on Long-Term Planning of Water Management," which endorse basin-wide, cooperative management of shared water resources.[132]

As things now stand, no solution is possible without the creation of the necessary new law: If a cooperative management system is to be put in place in the Iberian peninsula, it must entail the creation of a legal mechanism not only capable of resolving disputes, but also capable of providing for considerable active cooperation in the joint management of resources.[133] The hydrologic and managerial imbalances between Spain and Portugal, however, are likely to make such

---

128. *See* Bunge, *supra* note 122, at 673-76, 681-83; Zacker, *supra* note 122, at 251-61, 264-78; Linda M. Sheehan, Comment, *The EEC's Proposed Directive on Civil Liability for Damage Caused by Waste: Taking Over When Prevention Fails*, 18 ECOL. L.Q. 405 (1991); Felicia A. Wartnik, Comment, *Waste Liability and the European Economic Community: An Analysis of a Proposed Directive on Civil Liability for Damage Caused by Waste*, 2 COLO. J. INT'L ENVTL. L. & POL'Y 429 (1991).

129. Single European Act, *supra* note 125, art. 130s. Thus far, the power to define nonunanimous issues has been used solely with regard to the setting of technical standards. Sheehan, *supra* note 128, at 413 n.63.

130. *Two Decades of Co-Operation on Water*, Economic Comm'n for Europe, adopted 35th Sess., Decision B (XXXV), at 1, 3, U.N. Doc. ECE/ENVWA/2 (1988) [hereafter ECE].

131. Decision C (XXXIX), in ECE, *supra* note 130, at 12, 15.

132. ECE, *supra* note 130, at 39, 41.

133. *See generally* Utton, *supra* note 74, at §§ 49.03. 49.05, 49.06.

solutions difficult to attain unless Spain accepts the twin rules of equitable utilization and no appreciable harm.[134]

What the situation in the Iberian peninsula requires is the creation, by agreement between the interested states, of a formal legal regime to manage actively the water resources shared between them.[135] A formal legal regime would have to create a system of cooperative management in a structure capable of determining the facts of water use in each nation, to resolve disputes between the interested nations, to guide responses to unusual temporary shortfalls of water, to regulate long-term answers to the serious permanent shortages, and to enforce its decisions. The two nations, however, have thus far shown little interest in negotiating such an arrangement.

Portugal should break with its usual "soft" approach to Spain[136] and undertake to initiate such negotiations and attempt to persuade Spain that its self-interest would also be furthered by such arrangements. The fact that Spain is in clear violation of customary international law[137] provides some leverage in beginning the bargaining, yet by itself such a claim will gain Portugal little; Portugal's unreciprocated dependence on water flowing from Spain[138] prevents Portugal from being able to impose effective unilateral sanctions on Spain. The institutions of the European Community, even if only through publicizing the problem and bringing informal pressure on Spain, will probably provide the only effective means of bringing Spain to recognize and honor Portuguese rights.

---

134. See part III of this Article.

135. *See, e.g.*, Stephen McCaffrey, *The Law of International Watercourses: Ecocide or Ecomanagement?*, 59 REV. JUR. U.P.R. 1003, 1004 (1990).

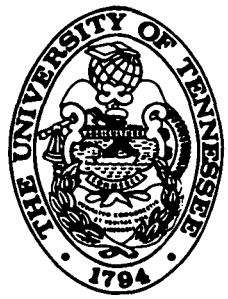
136. Evaristo da Silva, *supra* note 10, at 9.

137. See part III(C) of this Article.

138. See part I of this Article.

Annual Index

TENNESSEE LAW REVIEW



Volume 59

THE UNIVERSITY OF TENNESSEE
COLLEGE OF LAW

INDEX TO VOLUME 59

Tennessee Law Review Association, Inc.
1505 W. Cumberland Ave.
Knoxville, Tennessee 37996-1800

AUTHOR INDEX

SUBJECT INDEX

TABLE OF CASES

N

T

Y

Z