HARVARD JOURNAL
on
LEGISLATION

VOLUME 33, NUMBER 1 WINTER 1996

STATUTE

ARTICLES

NOTE

HARVARD JOURNAL ON LEGISLATION
Harvard Law School
Cambridge, MA 02138

Phone: (617) 495-4400
Internet: HLSJOL@HULAW1.HARVARD.EDU

The HARVARD JOURNAL ON LEGISLATION (ISSN 0017-808x) is published twice each year (winter and summer) by Harvard Law School students. Third-class postage paid at Boston, Massachusetts, and at additional mailing offices.

Publication Policy

The JOURNAL specializes in the analysis of legislation and the legislative process. It focuses on legislative reform and on organizational and procedural factors that affect the efficiency and effectiveness of legislative decision-making. The JOURNAL is especially interested in publishing articles that examine public policy problems of nationwide significance and propose legislation to resolve them.

The JOURNAL welcomes unsolicited manuscripts. Authors interested in being published in the JOURNAL should send their manuscripts and all related correspondence to: Articles Editor, HARVARD JOURNAL ON LEGISLATION, at the above address. Manuscripts should be typed, double-spaced, and submitted in duplicate.

Views expressed in the JOURNAL are those of the authors and do not necessarily reflect the views of the JOURNAL's officers, editors, or other members.

Subscriptions

The JOURNAL's subscription rate for the 1996 volume (Volume 33) is $24.00 domestic rate; $28.00 foreign surface rate ($30.00 in Canada); $42.00 foreign air mail. All correspondence concerning subscriptions should be addressed to the Subscriptions Manager, Publications Center, Harvard Law School, Cambridge, MA 02138.

Unless otherwise requested by the subscriber, the JOURNAL is mailed at bulk rate. For some overseas subscribers, delivery via air mail may be more reliable than standard surface rate service. Claims for missing issues of Volume 33 will be honored only until January 1997.

Individual Issues

Copies of individual issues from the current volume can be obtained directly from the JOURNAL for $12.00 each (plus $2.00 per copy for foreign mailing).

Copies of issues from all past volumes of the JOURNAL are available from Fred B. Rothman & Co., 10368 West Centennial Road, Littleton, CO 80123, (303) 979-5657. Contact Fred B. Rothman & Co. for information on prices for bound volumes.

When purchasing copies of issues from either the JOURNAL or Fred B. Rothman & Co., please send a check for the purchase price when you place your order.

Indexes

Starting with Volume 23, the JOURNAL discontinued publication of volume-by-volume and five-year cumulative indexes. The JOURNAL is indexed in the *Index to Legal Periodicals*.

STATUTE

A MODEL STATE ACT TO AUTHORIZE AND REGULATE PHYSICIAN-ASSISTED SUICIDE

CHARLES H. BARON
CLYDE BERGSTRESSER
DAN W. BROCK
GARRICK F. COLE
NANCY S. DORFMAN
JUDITH A. JOHNSON
LOWELL E. SCHNIPPER
JAMES VORENBERG
SIDNEY H. WANZER*

---

* Charles H. Baron, Professor of Law, Boston College Law School; A.B., University of Pennsylvania, 1958; LL.B., Harvard Law School, 1961; Ph.D. (Philosophy), University of Pennsylvania, 1972. Clyde Bergstresser, Member, Bergstresser & Associates, Boston; B.S., New Jersey Institute of Technology, 1971; J.D., Boston College Law School, 1974. Dan W. Brock, University Professor, Professor of Philosophy and Biomedical Ethics, and Director, Center for Biomedical Ethics, Brown University; B.A., Cornell University, 1960; Ph.D. (Philosophy), Columbia University, 1970. Garrick F. Cole, Member, Smith, Duggan and Johnson, Boston; B.A., Harvard College, 1968; J.D., Boston College Law School, 1973. Nancy S. Dorfman, Chairman, Greater Boston Hemlock Society, Belmont, Mass.; B.A., University of California, Berkeley, 1945; Ph.D. (Economics), University of California, Berkeley, 1967. Judith A. Johnson, Vice President for Legal Services, New England Medical Center Hospitals, Inc., Boston; Adjunct Assistant Professor of Community Health, Tufts University School of Medicine; B.A., Brown University, 1966; M.A., Temple University, 1970; J.D., Boston College Law School, 1981. Lowell E. Schnipper, Chief, Oncology Division, Beth Israel Hospital, Boston; Theodore W. and Evelyn G. Berenson Associate Professor of Medicine, Harvard Medical School; B.S., Queens College, 1966; M.D., Downstate College of Medicine, State University of New York, 1968. James Vorenberg, Roscoe Pound Professor of Law and former Dean, Harvard Law School; A.B., Harvard College, 1949; LL.B., Harvard Law School, 1951. Sidney H. Wanzer, Director, Harvard Law School Health Services; A.B., Duke University, 1950; M.D., Duke University School of Medicine, 1954.

Many of the ideas and opinions expressed in this Article represent the consensus of the authors. Others, however, represent the agreement of only a majority of the authors. As a result, the ideas and opinions expressed should not be attributed to any of the authors individually. Furthermore, they should not be attributed to any of the authors' respective employers.

Comments about this Article may be addressed via e-mail to Professor Baron at BARON@HERMES.BC.EDU.

Despite laws in many states prohibiting assisted suicide, an unknown but significant number of people each year commit suicide with the aid of a physician. In recent years, the phenomenon of physician-assisted suicide has attracted greater attention as physicians have openly risked prosecution to shed light on the subject, advocates have raised a series of legal challenges to laws banning assisted suicide, and a federal judge has struck down the nation's first statute allowing physicians to assist patients in suicide.

In this Article, nine authors from the fields of law, medicine, philosophy, and economics propose a comprehensive statute to permit and regulate physician-assisted suicide for patients suffering from terminal illnesses or unbearable pain. The proposed statute provides a specific series of procedural requirements designed to prevent mistaken decisions and affords limited legal protection to physicians who follow its requirements.

In recent years, the prerogatives of competent patients to make end-of-life medical treatment decisions have been clarified, afforded legal protection, and increasingly accepted in medical practice.[1] These prerogatives include the right of competent patients to hasten the moment of their death by refusing treatment that would otherwise prolong their suffering.[2] Under legal regimes that afford terminal patients this prerogative, physicians and other health care practitioners must comply with the decisions of such patients to withhold or withdraw medical treatment and may do so without fear of legal liability.[3] As rights to forgo life-sustaining treatment have become established at law, many people have come to believe that a patient's control over his or her dying should be extended to permit active means to hasten death when there is no life-sustaining treatment to forgo.[4]

---

[1] *See* COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS, AMERICAN MEDICAL ASS'N, CODE OF MEDICAL ETHICS: CURRENT OPINIONS WITH ANNOTATIONS § 2.20 (1994) [hereinafter CODE OF MEDICAL ETHICS] (recognizing that patient preferences with regard to life-prolonging treatment should prevail). *See generally* THE HASTINGS CENTER, GUIDELINES ON THE TERMINATION OF LIFE-SUSTAINING TREATMENT AND CARE OF THE DYING (1987); PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, DECIDING TO FOREGO LIFE-SUSTAINING TREATMENT (1983); ROBERT F. WEIR, ABATING TREATMENT WITH CRITICALLY ILL PATIENTS (1989); George J. Annas & Leonard H. Glantz, *The Right of Elderly Patients to Refuse Life-Sustaining Treatment*, 64 MILBANK Q. supp. no. 2, at 95 (1986); Sidney H. Wanzer et al., *The Physician's Responsibility Toward Hopelessly Ill Patients*, 310 NEW ENG. J. MED. 955 (1984).

[2] *See* UNIF. RIGHTS OF THE TERMINALLY ILL ACT § 2 (1989) (enacted in seven states and Virgin Islands, previous version enacted in six states); *id.* introductory comment, 9B U.L.A. supp. 135 (1995) (citing similar laws in 31 states and District of Columbia); Cruzan v. Director, Mo. Dep't of Health, 497 U.S. 261, 269–79 (1990) (suggesting the existence of a liberty interest in refusing such treatment and reviewing state and federal cases).

[3] *See* UNIF. RIGHTS OF THE TERMINALLY ILL ACT §§ 3, 9 (1989).

[4] *See, e.g.*, JAMES RACHELS, THE END OF LIFE: EUTHANASIA AND MORALITY (1986); HUMPHREY TAYLOR, DOCTOR-ASSISTED SUICIDE: SUPPORT FOR DR. KEVORKIAN REMAINS STRONG AND 2-TO-1 MAJORITY APPROVES OREGON-STYLE ASSISTED SUICIDE BILL (The Harris Poll No. 9, 1995); Marcia Angell, *Euthanasia*, 319 NEW ENG. J. MED.

The issue remains a source of ethical, religious, and legal controversy. Anecdotal reports and occasional confidential surveys of physicians reveal that some physicians occasionally assist patients with suicide,[5] but data on the frequency with which physician-assisted suicide occurs are not reliable.[6] Moreover, threats of criminal charges and civil litigation make even the most empathetic physicians wary of complying with a patient's request for such assistance in the absence of clear-cut legal guidance and protection.[7]

Sharing the belief that physician-assisted suicide should be an option available to competent patients, we met together over a two-year period to draft a model statute to authorize physician-assisted suicide. Several of us were panel members at a symposium sponsored by the Massachusetts Bar Association in 1992 that focused on the state of the law in the Commonwealth concerning assistance in dying.[8] With the addition of several others, we authors now include three attorneys who represent patients, hospitals, and physicians; two law professors with interests in medical and constitutional law; a professor of philosophy who specializes in bioethics; a patient advocate and public policy economist; and two physicians with experience in academic medicine and community practice.

1348 (1988); Dan W. Brock, *Voluntary Active Euthanasia*, HASTINGS CENTER REP., Mar.-Apr. 1992, at 10 [hereinafter Brock, *Voluntary Active Euthanasia*]; Howard Brody, *Assisted Death—A Compassionate Response to a Medical Failure*, 327 NEW ENG. J. MED. 1384 (1992); Christine K. Cassel & Diane E. Meier, *Morals and Moralism in the Debate over Euthanasia and Assisted Suicide*, 323 NEW ENG. J. MED. 750 (1990); Note, *Physician-Assisted Suicide and the Right to Die with Assistance*, 105 HARV. L. REV. 2021 (1992); Richard A. Knox, *Poll: Americans Favor Mercy Killing*, BOSTON GLOBE, Nov. 3, 1991, at 1.

[5] *See, e.g.*, Robyn S. Shapiro et al., *Willingness to Perform Euthanasia: A Survey of Physician Attitudes*, 154 ARCHIVES INTERNAL MED. 575, 581 (1993) (revealing that 2.2% of physicians surveyed had performed euthanasia); Timothy E. Quill, *Death and Dignity: A Case of Individualized Decision Making*, 324 NEW ENG. J. MED. 691 (1991) (firsthand account by physician of assisted suicide); Dick Lehr, *Death & the Doctor's Hand: Increasingly, Secretly, Doctors Are Helping the Incurably Ill to Die*, BOSTON SUNDAY GLOBE, Apr. 25, 1993, at 1 (profiling two doctors who have assisted patients in suicide); New Hampshire Medical Society, End-of-Life Issues: Survey Results (Sept. 17, 1994) (press release, on file with the *Harvard Journal on Legislation*) (reporting that 4.4% of physicians responding had prescribed a lethal dose of medication for a terminally ill patient and that 1.9% had administered a lethal dose to such a patient).

[6] *See* Shapiro et al., *supra* note 5, at 576 (noting 33% response rate); New Hampshire Medical Society, *supra* note 5, at 2 (noting 44% response rate).

[7] *See infra* part I.B.

[8] For further information on the symposium, see Massachusetts Bar Ass'n, Assisted Suicide & the Right to Die: A Massachusetts Perspective (Nov. 1992) (symposium materials, on file with the *Harvard Journal on Legislation*).

Part I of this Article explains the relationship of physician-assisted suicide to the current law and to current thinking in medicine and philosophy. Part II explores the difficult choices that we made in determining what form of physician-assisted suicide should be available, who should be able to receive assistance, and how simultaneously to protect privacy and prevent abuse. Part III examines the constitutionality of our model statute. Finally, Part IV presents a detailed overview of the provisions of our statute.

I. The Medical, Ethical, and Legal Context of Physician-Assisted Suicide

The statute that we propose is designed to provide the option of physician-assisted suicide to competent patients who either have a terminal illness or are suffering from unrelievable and unbearable distress, due to bodily illness, that is so great that they prefer death.[9] The statute can be fully understood only in light of current medical, ethical, and legal constraints on physician-assisted suicide.

A. *The Medical and Moral Basis for Physician-Assisted Suicide*

We believe that it is reasonable to provide relief from suffering for patients who are dying or whose suffering is so severe that it is beyond their capacity to bear. Some opponents of physician-assisted suicide see such a step as a radical moral departure from present medical practice,[10] but we believe it is consistent with the fundamental values underlying the legal and ethical requirements of respect for the right of competent patients to give or withhold their consent to any treatment, including life-sustaining treatment.[11] The most basic values that support and guide all

---

[9] *See infra* A Model State Act to Authorize and Regulate Physician-Assisted Suicide § 1 [hereinafter Model Act].

[10] *See, e.g.*, Willard Gaylin et al., *Doctors Must Not Kill*, 259 JAMA 2139 (1988) (opposing assisted suicide as inconsistent with medical principles); Leon R. Kass, *Neither for Love nor Money: Why Doctors Must Not Kill*, Pub. Interest, Winter 1989, at 25 (opposing assisted suicide as, *inter alia*, unprofessional, harmful to doctor-patient relationship, and a violation of Hippocratic Oath).

[11] *See* Cobbs v. Grant, 502 P.2d 1, 9 (Cal. 1972); Harnish v. Children's Hosp. Medical Ctr., 439 N.E.2d 240, 242 (Mass. 1982); Unif. Rights of the Terminally Ill Act

health care decision making, including decisions about life-sustaining treatment, are the same values that provide the fundamental basis for physician-assisted suicide: promoting patients' well-being and respecting their self-determination or autonomy.[12]

The legal right to decide about life-sustaining treatment has given most patients appropriate control over their own dying, and we believe strongly that this control, along with proper supportive care, meticulous attention to details, and truly adequate pain relief measures, will meet the needs of the great majority of dying patients and usually obviate the occasion for the patient to consider the possibility of hastening death.[13] However, for some patients who are undergoing severe suffering and confronting an unbearable or meaningless existence, either no life-sustaining treatment is available to be forgone or forgoing such treatment will result in a prolonged, unbearable, and inhumane dying process. Even when optimal care has been given, intolerable distress may remain in these patients, such that they may conclude rationally that hastening death is the only appropriate goal.[14] For these patients, more active means of hastening death are necessary, supported by the very same values that promote patients' well-being and respect their self-determination.

Viewed in this way, making physician-assisted suicide available to patients who choose it is not a radical departure in medical practice or public policy, but a natural and appropriate extension of presently accepted practices. Physicians are uniquely able to provide this necessary assistance with a combination of expert knowledge, compassionate concern for the patient, pro-

---

§ 2 (1989); DAN W. BROCK, *Death and Dying, in* LIFE AND DEATH: PHILOSOPHICAL ESSAYS IN BIOMEDICAL ETHICS 144, 148–53 (1993).

[12] *See generally* PRESIDENT'S COMM'N FOR THE STUDY OF ETHICAL PROBLEMS IN MEDICINE AND BIOMEDICAL AND BEHAVIORAL RESEARCH, MAKING HEALTH CARE DECISIONS (1982).

[13] *But see* Marcia Angell, *The Quality of Mercy*, 306 NEW ENG. J. MED. 98 (1982) (calling for renewed attention to the problem of inadequate pain relief); Charles S. Cleeland et al., *Pain and Its Treatment in Outpatients with Metastatic Cancer*, 330 NEW ENG. J. MED. 592 (1994) (noting that many cancer patients receive inadequate pain treatment); Marilee M. Donovan et al., *Incidence and Characteristics of Pain in a Sample of Medical-Surgical Inpatients*, 30 PAIN 69 (1987) (recognizing that treatment of pain remains a significant problem); Robert D. Truog et al., *Barbiturates in the Care of the Terminally Ill*, 327 NEW ENG. J. MED. 1678 (1992) (noting tension between easing pain and hastening death).

[14] *See* TIMOTHY E. QUILL, DEATH AND DIGNITY: MAKING CHOICES AND TAKING CHARGE 104–13 (1993); Sidney H. Wanzer et al., *The Physician's Responsibility Toward Hopelessly Ill Patients: A Second Look*, 320 NEW ENG. J. MED. 844, 847–48 (1989).

fessional responsibility to the patient and to society, and the ability to determine and prescribe the medication that the patient will usually require to achieve a humane and certain death.[15] They should be able lawfully to provide the assistance necessary to achieve that goal. Our model statute would allow such assistance, while at the same time attempting to provide adequate protection against possible abuses.

B. *Current Legal Obstacles to Physician-Assisted Suicide*

In a jurisdiction without a statute authorizing physician-assisted suicide, a physician who provided means of suicide to a patient could be convicted of manslaughter[16] or a specific crime of aiding or assisting a suicide or an attempted suicide.[17] Under certain circumstances, such a physician could be convicted of murder, but in many states, a murder conviction requires active participation in the death rather than merely supplying the means of death.[18] Nevertheless, even the possibility of murder charges is likely to have a deterrent effect on a physician who would otherwise consider assisting a patient to commit suicide. Indeed, even in a jurisdiction where assisted suicide is not prohibited by statute, a physician who assisted in a patient's suicide could be convicted of a common-law felony.[19]

Among the civil threats to physicians undertaking assisted suicide are liability for wrongful death[20] and medical malpractice.[21] A physician might also face professional sanctions, either

---

[15] *See* Ann Alpers & Bernard Lo, *Physician-Assisted Suicide in Oregon: A Bold Experiment*, 274 JAMA 483 (1995) (suggesting a number of issues to be considered by physicians in light of legalization of physician-assisted suicide).

[16] *See, e.g.*, N.Y. PENAL LAW § 125.15(3) (McKinney 1987).

[17] *See, e.g.*, MODEL PENAL CODE § 210.5(2) (1962); N.Y. PENAL LAW § 120.30 (McKinney 1987).

[18] *See, e.g.*, People v. Cleaves, 280 Cal. Rptr. 146, 151 (Cal. Ct. App. 1991); People v. Kevorkian, 527 N.W.2d 714, 738–39 (Mich. 1994), *cert. denied*, 115 S. Ct. 1795 (1995); State v. Sexson, 869 P.2d 301, 304 (N.M. Ct. App. 1994).

[19] *See, e.g.*, *Kevorkian*, 527 N.W.2d at 739.

[20] *See, e.g.*, MASS. GEN. L. ch. 229, § 2 (1994); 42 PA. CONS. STAT. ANN. § 8301(a) (Supp. 1995).

[21] A physician can be found liable for malpractice when a patient commits suicide against the wishes of the physician. *See, e.g.*, Peoples Bank of Bloomington v. Damera, 581 N.E.2d 426, 429 (Ill. App. Ct. 1991); Stepakoff v. Kantar, 473 N.E.2d 1131, 1135 (Mass. 1985); Champagne v. United States, 513 N.W.2d 75, 76–77 (N.D. 1994). By the same reasoning, a physician who actually *expected* a patient to commit suicide could be found liable.

as a result of specific ethical prohibitions on assisted suicide[22] or because of the philosophical or political opposition of the reviewing disciplinary board. Finally, a physician who assisted in a suicide could lose staff privileges at a hospital that objected to the practice.

The net result of these obstacles to physician-assisted suicide is to deter physicians from considering the practice, even if they might otherwise have no objection to it.[23] As we explain in the next section, we believe that a statute is needed to enable physicians to assist patients in suicide in appropriate circumstances.

C. *The Need for a Specific Statute*

Laws that deprive persons of access to physician-assisted suicide have been challenged recently on constitutional grounds in federal and state courts in several jurisdictions.[24] We feel that a preferable way to establish a right to physician-assisted suicide is to make this option available to persons through explicit statutory authorization. Even if laws restricting assisted suicide are struck down, laws or regulations will be necessary to provide oversight and protection against abuse.[25] Our statutory approach permits the careful development of procedures necessary to limit abuse. A statute also more clearly requires and establishes the public support that should exist for the practice before it is made legally available.

---

[22] *See* CODE OF MEDICAL ETHICS, *supra* note 1, § 2.211 ("Physician assisted suicide is fundamentally incompatible with the physician's role as healer . . . ."). The Hippocratic Oath also prohibits direct assistance in death. *See* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 226–27 (4th ed. 1994).

[23] *See* Shapiro et al., *supra* note 5, at 581 (noting that although 35.2% of physicians responding had been asked to perform euthanasia and 27.8% would be willing to perform euthanasia if it were legal, only 2.2% had actually performed it).

[24] *See* Compassion in Dying v. Washington, 49 F.3d 586 (9th Cir.), *reh'g en banc granted*, 62 F.3d 299 (9th Cir. 1995); Quill v. Koppell, 870 F. Supp. 78 (S.D.N.Y. 1994); People v. Kevorkian, 527 N.W.2d 714 (Mich. 1994), *cert. denied*, 115 S. Ct. 1795 (1995). *See generally* Yale Kamisar, *Are Laws Against Assisted Suicide Constitutional?*, HASTINGS CENTER REP., May-June 1993, at 32 (arguing against a right to assisted suicide).

[25] *See* Guy I. Benrubi, *Euthanasia—The Need for Procedural Safeguards*, 326 NEW ENG. J. MED. 197 (1992); Franklin G. Miller et al., *Regulating Physician-Assisted Death*, 331 NEW ENG. J. MED. 119 (1994); Timothy E. Quill et al., *Care of the Hopelessly Ill: Proposed Clinical Criteria for Physician-Assisted Suicide*, 327 NEW ENG. J. MED. 1380 (1992).

Commentators have argued that there is no need for legislation in states where assisted suicide is not specifically outlawed by statute, because physicians in those states may legally provide patients with means of suicide or, in any event, need not fear prosecution for doing so.[26] Others have maintained that to legalize physician-assisted suicide would make suicide "too easy," opening the option to patients whose conditions do not warrant such an extreme measure and risking that it would be urged on patients who do not want it.[27] Some contend that legislation would impose onerous regulations on the conduct of a procedure that already takes place when, in the judgment of the physician, the situation warrants it.[28]

On the contrary, for the following reasons, we believe that society and the medical profession would be better served by a statute that expressly permits physician-assisted suicide under certain well-defined circumstances, rather than by no law at all:

First, in states that do not explicitly prohibit any form of assisted suicide, the law's silence leaves physicians in serious doubt concerning the legality of providing means of suicide to a patient,[29] while in states that do outlaw assisted suicide, physicians must risk prosecution for a felony in order to assist in a patient's suicide.[30] As a result, patients who seek means of dying are often denied assistance,[31] and success in finding a physician who will help may be a result of luck more than of need.

Second, physicians who now provide assistance in suicide may be compelled by fear of prosecution to do so in secret,[32] without the opportunity to discuss the case fully and freely with colleagues or other professionals. In contrast, physicians have ac-

---

[26] *See, e.g.*, Leonard H. Glantz, *Withholding and Withdrawing Treatment: The Role of the Criminal Law*, 15 LAW, MED. & HEALTH CARE 231, 232 (1987–1988) ("No physician has ever been successfully prosecuted for an act of either omission or commission that led to the death of a seriously ill patient."). As noted above, however, the lack of a statute prohibiting physician-assisted suicide does not preclude prosecution. *See supra* note 19 and accompanying text.

[27] *See, e.g.*, George J. Annas, *Death by Prescription—The Oregon Initiative*, 331 NEW ENG. J. MED. 1240, 1243 (1994) (noting risks to poor, elderly, and minorities); J. David Velleman, *Against the Right to Die*, 17 J. MED. & PHIL. 665, 675 (1992) (recognizing danger of coercion).

[28] *See, e.g.*, Annas, *supra* note 27, at 1242–43.

[29] *See supra* note 19 and accompanying text.

[30] *See supra* text accompanying notes 16–18.

[31] *See* DEREK HUMPHRY, LET ME DIE BEFORE I WAKE 7–11, 34–44 (5th ed. 1987) [hereinafter HUMPHRY, LET ME DIE] (relating stories of two patients whose physicians refused to aid them in suicide); Shapiro et al., *supra* note 5, at 581.

[32] *But see* Timothy E. Quill, *Death and Dignity: A Case of Individualized Decision Making*, 324 NEW ENG. J. MED. 691 (1991).

cess to a variety of professional consultations, often including review by ethics committees or consultants, in connection with other profoundly serious medical-ethical decisions.[33]

Third, physicians who now provide assistance in suicide do so without any form of accountability, procedures, requirements, or guidelines to assure that the patient's request for assistance is competent, fully informed, voluntary, and enduring and that the diagnosis and treatment options have been confirmed and fully explained to the patient.

Fourth, in the absence of assistance from a physician, many terminally ill patients now attempt to end their lives on their own, often in ignorance of and without access to the best means of doing so.[34]

Fifth, some terminally ill patients prematurely elect to end their lives by forgoing treatment because they fear that the opportunity to end their lives will not arise later should their suffering become unendurable.[35]

Finally, with or without assistance from a physician, many patients who end their lives may feel obliged to do so in solitude, without the professional advice of a physician or the presence and comfort of loved ones.

II. THREE FUNDAMENTAL ISSUES

A. *Active Euthanasia Versus Physician-Assisted Suicide*

Our proposed statute would legalize physician-assisted suicide under certain conditions, but it does not address voluntary active euthanasia. By "physician-assisted suicide," we mean providing

---

[33] *See* Troyen A. Brennan, *Ethics Committees and Decisions to Limit Care: The Experience at the Massachusetts General Hospital*, 260 JAMA 803 (1988); John LaPuma et al., *An Ethics Consultation Service in a Teaching Hospital: Utilization and Evaluation*, 260 JAMA 808 (1988). *See generally* INSTITUTIONAL ETHICS COMMITTEES AND HEALTH CARE DECISION MAKING (Ronald E. Cranford & A. Edward Doudera eds., 1984).

[34] *See, e.g.*, GEORGE H. COLT, THE ENIGMA OF SUICIDE 373 (1991) (reporting several disastrous suicide attempts); HUMPHRY, LET ME DIE, *supra* note 31, at 45–55 (relating story of a bungled suicide attempt); Jody B. Gabel, *Release from Terminal Suffering?: The Impact of AIDS on Medically Assisted Suicide Legislation*, 22 FLA. ST. U. L. REV. 369, 384–95 (1994) (discussing a nearly botched suicide).

[35] *See* DEREK HUMPHRY, FINAL EXIT: THE PRACTICALITIES OF SELF-DELIVERANCE AND ASSISTED SUICIDE FOR THE DYING 103–05 (1991); Stephen A. Newman, *Euthanasia: Orchestrating "The Last Syllable of . . . Time"*, 53 U. PITT. L. REV. 153, 183 (1991).

the patient with the means, such as a drug that can be lethal in certain doses, to end his or her own life. Voluntary active euthanasia, in contrast, requires the active participation of the physician in performing the action, such as administering a lethal injection, that ends the patient's life. Members of the public and the medical community disagree, and we disagree among ourselves, as to whether there is an important difference between the two concepts.[36]

We have chosen to allow only physician-assisted suicide for two main reasons. First, we consider the voluntariness of the patient's act to be critical. Restricting the statute to physician-assisted suicide provides in many cases a stronger assurance of the patient's voluntary resolve to die and of the central role of patient responsibility for the act. Second, we believe that there would be greater acceptance of the model statute by the public, legislators, and physicians if it were limited to physician-assisted suicide, partly because of the public perception of voluntariness and partly because of the strong ethical objections of some physicians and others to euthanasia.[37]

B. *Which Patients Should be Eligible for Physician-Assisted Suicide?*

We agreed from the outset that to be eligible for physician-assisted suicide, the patient must be an adult, aged eighteen years or older.[38] We also agreed that anyone who is terminally ill, that is, likely to die from an illness within six months, should qualify without having to demonstrate that his or her suffering is unbearable.[39] We continued to debate until the very end of our

---

[36] *Compare, e.g.*, Diane E. Meier, *Physician-Assisted Dying: Theory and Reality*, 3 J. CLINICAL ETHICS 35, 35 (1992) (significant difference between the two) *with, e.g.*, Brody, *supra* note 4, at 1386 (a psychological, but not an ethical, difference) *and* Brock, *Voluntary Active Euthanasia*, *supra* note 4, at 10 (no significant difference).

[37] *See supra* note 10 and accompanying text.

[38] *See* Model Act § 3(a)(1).

[39] *See id.* § 2(*i*). Patients with terminal illnesses have generally been seen as the least controversial candidates for the recognition of a right to die. Early decisions in this field began by recognizing the right of such patients to refuse life-prolonging treatment. *See, e.g.*, Superintendent of Belchertown State School v. Saikewicz, 370 N.E.2d 417 (Mass. 1977). Likewise, early living-will legislation offered the right to refuse life-prolonging treatment only to those with terminal illnesses. *See, e.g.*, California Natural Death Act, sec. 1, § 7187(e)–(f), 1976 Cal. Stat. 6478, 6479 (repealed 1991). The fact that terminal patients will die soon, with or without treatment, may be seen as reducing the strength of any countervailing state interest in preventing such patients from deciding to shorten their lives further and as reducing the cost of any errors that may

deliberations as to how far, if at all, to broaden this eligibility beyond the six-month limit. Our major concern was whether and how to extend the option to patients who are not likely to die from their illnesses within six months but have bodily disorders that cause intractable and unbearable suffering, such as AIDS, advanced emphysema, some forms of cancer, amyotrophic lateral sclerosis, multiple sclerosis, and many other debilitating conditions.

With respect to this issue, we faced the difficulty of defining unbearable suffering in a sufficiently objective fashion that physician-assisted suicide would not be available to everyone who had some form of physical or psychological suffering and merely requested it. In the end, a bare majority of us agreed to allow anyone to be eligible whose illness is incurable and who subjectively feels that the accompanying suffering is worse than death.[40] We rejected a more objective definition of the patient's suffering for two principal reasons. First, we found that it was not possible to construct an objective definition that was not overly restrictive as to the patients who would meet it. Second, and more important, we realized that whether one's suffering is sufficiently unbearable to make death preferable to continued life is an inherently subjective determination on which people differ, and for which no objective standard should be imposed on everyone. Because the statute does not endow the patient with a right to physician-assisted suicide, however, the physician still retains the ability to decide whether the case warrants providing such relief. In addition, because the statute requires competency,[41] the subjective preference for death of a clinically depressed or mentally ill patient would be insufficient to qualify that patient for assisted suicide.

---

be made in the process of the decision to refuse treatment. The physical and psychological pain suffered by a terminally ill patient also suggests that his or her desire to hasten death may be reasonable. Finally, the restriction of the right to the terminally ill establishes a boundary that helps to address slippery-slope concerns. *See infra* text accompanying note 42.

[40] *See* Model Act § 2(d).

[41] *See id.* § 3(a)(3)(A).

C. *Protecting Patients and Physicians Versus Maintaining Privacy*

Procedural safeguards that adequately protect both patients and physicians unavoidably conflict with the privacy of patients and families and the privacy of the physician-patient relationship. To maximize privacy, we considered proposing a statute that would simply state in very general terms that physician-assisted suicide was legal under certain stated factual circumstances but would not prescribe procedural requirements. Under this abbreviated approach, an assisted-suicide statute might comprise only a few simple provisions to the effect that a physician would not be guilty of unlawfully assisting a patient to commit suicide, provided that: (1) the physician's assistance were limited to making available a substance used by the patient to end the patient's life; (2) the patient had an illness that was either terminal or caused the patient intractable and unbearable suffering; (3) the patient had made a decision to hasten death because of the illness; and (4) the patient's decision was fully informed as to relevant medical facts and was not the result of a mental illness or undue influence from other persons. We concluded that such an abbreviated approach would not adequately protect patients or physicians.

The procedures, conditions, and documentation requirements built into the model statute are designed to ensure that physician-assisted suicide is restricted to patients who are truly terminally ill or suffering from intractable and unbearable illnesses, and whose requests are demonstrably competent, fully informed, voluntary, and enduring. To govern the practice in accordance with these principles, it is necessary that the statute contain strong safeguards and precise procedural requirements. Such detailed requirements will counter a common objection to making physician-assisted suicide legally permissible: the so-called "slippery slope" argument.[42] While it is not possible to guarantee that abuse and unjustified extension of the practice cannot or will not take place, we believe strong and effective safeguards, together with a clear understanding of the rationale for the practice and

---

[42] *See, e.g.*, Daniel Callahan, *When Self-Determination Runs Amok*, HASTINGS CENTER REP., Mar.-Apr. 1992, at 52, 54; Gaylin et al., *supra* note 10, at 2139–40; Peter A. Singer & Mark Siegler, *Euthanasia—A Critique*, 322 NEW ENG. J. MED. 1881, 1883 (1990). *See generally* Wibren van der Burg, *The Slippery Slope Argument*, 102 ETHICS 42 (1991) (outlining various forms of the argument).

the limits to which it applies, can reasonably meet concerns about a slippery slope.[43]

From the physician's perspective, an abbreviated approach such as that described above would preserve the physician's autonomy, would avoid imposing burdensome regulations on the physician, and would not intrude into the physician-patient relationship. It would not, however, adequately protect physicians and could make them unwilling to provide assistance in suicide even in appropriate situations. Because the conditions under which physicians could legally assist patients in suicide would be stated so generally, physicians would not know in advance whether a particular case fit those conditions and what actions they should take to obviate any significant risk of criminal charges. Even if a physician acted on a good-faith belief that the statutory conditions were met, he or she might be vulnerable to legal charges later. This possibility would almost certainly leave many physicians, who might have no principled objection to physician-assisted suicide, reluctant to provide it to any of their patients who might request it.[44]

Thus, not only for the protection of patients, but also for the protection of physicians, we chose to outline specific requirements that, when followed, offer the physician legal protection. Moreover, we concluded that extensive safeguards would both protect the integrity of the medical profession and help ensure that public trust in that integrity remains warranted.[45] If the public is to ask the medical profession to participate in physician-assisted suicide, then strong safeguards are a reasonable cost for the public and patients to bear.

It would be a mistake, however, to think that procedural safeguards do not come at a significant cost to the patient and to the physician-patient relationship. At what will typically be an emotionally difficult time for the patient and family, unfamiliar third-party consultants, evaluators, and witnesses must intrude into the physician-patient relationship. Patients and their families will often quite reasonably view the procedures as a profound inva-

---

[43] *Cf.* Margaret Battin, *Voluntary Euthanasia and the Risks of Abuse: Can We Learn Anything from the Netherlands?*, 20 LAW, MED. & HEALTH CARE 134 (1992) (advocating voluntary euthanasia if accompanied by strong procedural safeguards).

[44] In this respect, a general statute would be little better than no statute at all. *See supra* note 23 and accompanying text.

[45] *See* Gaylin et al., *supra* note 10, at 2139–40; David Orenthlicher, *Physician Participation in Assisted Suicide*, 262 JAMA 1844 (1989).

sion of their privacy at a point when time is short and privacy is especially important. We feel, nevertheless, that such procedures are necessary in order to ensure that in less-than-ideal relationships and conditions, misuse or abuse of the practice of physician-assisted suicide does not occur.

The detailed procedures also provide an openness to the practice of physician-assisted suicide that can give society greater assurance that the practice is operating as intended, and can provide feedback to government and professional bodies about needed refinements and revisions in the practice over time. In our final formulation of the statute, we therefore leaned in the direction of more extensive and comprehensive safeguards, acknowledging the costs to some patients and physicians.

III. CONSTITUTIONALITY OF THE MODEL ACT: *LEE V. OREGON*

In November 1994, Oregon voters enacted by initiative the nation's first statute explicitly permitting and regulating physician-assisted suicide.[46] The Oregon Act, which is similar in a number of respects to our proposed statute,[47] was promptly challenged in federal court on grounds that it violated the Fourteenth Amendment to the United States Constitution. On August 3, 1995, in *Lee v. Oregon*,[48] District Judge Michael R. Hogan declared the statute unconstitutional under the Equal Protection Clause of the Fourteenth Amendment.[49] The case is now on appeal to the United States Court of Appeals for the Ninth Circuit.[50]

We believe that the *Lee* case was wrongly decided and that our proposed statute will withstand appropriate constitutional scrutiny.

In 1990, the Supreme Court of the United States rendered its only decision to date on the subject of the right to die. In *Cruzan v. Director, Missouri Department of Health*,[51] the Court held that, where an incompetent patient is involved, a state may constitutionally require "clear and convincing" proof that the patient

---

[46] Oregon Death With Dignity Act, 1995 Or. Laws ch. 3 [hereinafter Oregon Act].

[47] In the interest of disclosure, we feel that we should mention that one of us played a minor role as an adviser to the drafters of the Oregon Act.

[48] 891 F. Supp. 1429 (D. Or. 1995).

[49] *Id.* at 1439.

[50] Lee v. Harcleroad, *appeal docketed*, No. 95-35804 (9th Cir. Aug. 11, 1995).

[51] 497 U.S. 261 (1990).

would want life-prolonging treatment withdrawn.[52] In passing, the Court recognized a patient's "constitutionally protected liberty interest" in refusing unwanted medical treatment.[53] Four Justices believed this liberty interest to be so strong in the context of a patient in a persistent vegetative state that they would have held the Missouri law restricting it unconstitutional under the Due Process Clause of the Fourteenth Amendment.[54] The other Justices, with the exception of Justice Scalia, also recognized such a constitutionally based right.[55] But they recognized as well a strong interest of the state in protecting the autonomy of an incompetent patient,[56] and they held that a state could constitutionally advance its interests, if it chose to do so, by requiring clear and convincing evidence of the patient's wishes.[57]

Judge Hogan's opinion in *Lee* turned *Cruzan* on its head. Whereas *Cruzan* dealt with state legislation that restricted a patient's right to be free from unwanted treatment, *Lee* dealt with state legislation advancing that right. Whereas *Cruzan* protected the right of the legislature to regulate the details of practice in this developing area, even though the regulation impinged upon a protected liberty interest, *Lee* struck down a popularly mandated measure that advanced that liberty interest.

The fault in the Oregon Act, from Judge Hogan's point of view, was that it did not advance patients' liberty interests as rationally as it might. In particular, the *Lee* court was concerned that (1) the Oregon Act permits "physicians who may not be psychiatrists, psychologists, or counselors to make an evaluation whether a condition is causing [the patient to exercise] impaired judgment";[58] (2) "[t]here is no requirement that the [patient] consult a certified social worker or other specialist to explore social services which might assist the person to live in greater comfort";[59] and (3) these and other failures in protection of the rights of patients apply only to the "terminally ill."[60] The court's sug-

---

[52] *Id.* at 280.
[53] *Id.* at 278.
[54] *See id.* at 316 (Brennan, J., dissenting, joined by Marshall and Blackmun, JJ.); *id.* at 350–51 (Stevens, J., dissenting).
[55] *See id.* at 278–79 (Rehnquist, C.J., for the Court); *id.* at 287 (O'Connor, J., concurring). *But see id.* at 299–300 (Scalia, J., concurring) (arguing that case does not implicate Constitution).
[56] *Id.* at 281–82.
[57] *Id.* at 284–85.
[58] *Lee*, 891 F. Supp. at 1435.
[59] *Id.*
[60] *Id.* at 1437.

gestion was that somehow the Oregon Act discriminated against the terminally ill as a class in violation of the Equal Protection Clause. Yet in the case of the Oregon Act and other legislation classifying patients on the basis of terminal illness, it is those persons who fear that they will one day find themselves among the terminally ill who are urging the enactment of such legislation to protect themselves from a lingering, undignified death. Clearly invidious motives are not at work when such statutes use terminal illness as a basis for classification.

Because the Oregon Act does not impinge upon a fundamental right and does not establish a classification on a basis that raises suspicions of invidious discrimination, the court was required to review the measure under the most lenient of constitutional standards.[61] It could find the Oregon Act unconstitutional only if one could conceive of no rational basis upon which the state could have used the means employed to advance a legitimate state interest.[62] In fact, Judge Hogan appears to have applied his own version of rational review and struck down the Oregon Act because it was not as rational as he thought it should have been. This sort of constitutional review is reminiscent of the discredited doctrine of *Lochner v. New York*.[63] A proper application of the rational-basis test would find both the Oregon Act and the statute that we propose here to be constitutional under the Fourteenth Amendment.

While we believe *Lee* will be reversed by the Ninth Circuit, we should note that our proposed statute addresses several of what Judge Hogan perceived to be the shortcomings of the Oregon Act. Our statute provides for a review of the patient's competency by a licensed psychiatrist, clinical psychologist, or psychiatric social worker;[64] allows patients the opportunity to consult with a social worker about alternatives to suicide;[65] and refuses to relieve physicians from liability for such actions as a negligent diagnosis.[66]

---

[61] *See* Schweiker v. Wilson, 450 U.S. 221, 230 (1981).
[62] *See id.*; FCC v. Beach Communications, Inc., 113 S. Ct. 2096, 2101 (1993).
[63] 198 U.S. 45 (1905).
[64] *See* Model Act § 5(b).
[65] *See id.* § 4(b).
[66] *See id.* § 13(c).

IV. Overview of the Model Act

A. *Who May Provide Physician-Assisted Suicide?*

The model statute allows a "responsible physician" to practice physician-assisted suicide and places a series of responsibilities on that physician.[67] The first question that we faced was who should be allowed to assume that role. Ideally, the physician who assists in a patient's suicide will be the one who has managed the patient's illness and who has a close professional relationship with the patient. However, the statute recognizes that because ethical constraints may prevent some physicians from assisting in suicide, a patient may need to have another physician provide him or her with the means of suicide. Section 2(h)[68] therefore allows any physician who has assumed full or partial responsibility for a patient's care to assume the role of responsible physician, even though he or she is not the patient's primary physician.

B. *Other Definitions*

Section 2(e) defines "medical means of suicide" as a medical substance or device prescribed for or supplied to a patient by the responsible physician. The use of the term "medical" requires that the means of suicide be otherwise consistent with sound medical practice; thus, providing a patient with an unapproved drug or a firearm (to take an extreme example) would not be permissible.

The definitions of "intractable and unbearable illness" and "terminal illness" are discussed above.[69] The remaining definitions in section 2 are self-explanatory.

C. *Conditions to be Met Before a Patient Receives Assistance in Suicide*

A fundamental goal of the statute is to protect patients from coercion or premature judgment. Section 3(a)(3) thus requires

---

[67] *See id.* § 3(a).

[68] All further references in the text of this Article to "section" are references to sections of the Model Act unless otherwise specified.

[69] *See supra* part II.B.

that four basic conditions be met before a physician may grant a patient's request for assisted suicide: the request must be competent, fully informed, voluntary, and enduring. The first three requirements are similar to those required for informed consent to ordinary medical treatment,[70] and the fourth is designed to ensure the consistent resolve of the patient. However, because of the seriousness and finality of the patient's decision, the requirements of the statute exceed those of consent to ordinary treatment.[71]

A competent request within the meaning of section 3(a)(3)(A) is a reasoned request for physician-assisted suicide from a patient, based on the patient's ability to understand his or her condition and prognosis, the benefits and burdens of available alternative treatments, and the consequences of suicide. A request distorted by clinical depression or other mental illness or impairment is not competent. However, the statute does not prohibit physician-assisted suicide for a patient suffering from clinical depression if the patient's judgment is not distorted—in other words, if the patient can make a reasoned decision consistent with his or her long-term values. A terminal illness is inherently depressing, and denying a patient assistance in suicide only because he or she feels sad or depressed would not be proper.[72] Nevertheless, the statute mandates that a professional mental health care provider evaluate the patient to determine that his or her decision is fully informed, free of undue influence, and not distorted by depression or any other form of mental illness.[73]

A fully informed request within the meaning of section 3(a)(3)(B) means that the patient understands the medical options available and their consequences. Section 4 requires the physician to dis-

---

[70] *See* Canterbury v. Spence, 464 F.2d 772, 782–89 (D.C. Cir.), *cert. denied*, 409 U.S. 1064 (1972); Cobbs v. Grant, 502 P.2d 1, 10–11 (Cal. 1972); Harnish v. Children's Hosp. Medical Ctr., 439 N.E.2d 240, 243–44 (Mass. 1982).

[71] Informed consent to ordinary treatment does not generally require consultations with respect to diagnosis or competency, witnessing of the informed-consent discussion, or documentation with the specificity required by our statute. *Compare* cases cited *supra* note 70 *with* Model Act §§ 4(d), 5.

[72] *See* Linda Ganzini et al., *The Effect of Depression Treatment on Elderly Patients' Preferences for Life-Sustaining Medical Therapy*, 151 AM. J. PSYCHIATRY 1631 (1994) (noting that in study of mild to moderate depression, remission of depression did not alter patients' desire for life-sustaining therapy); Melinda A. Lee, *Depression and Refusal of Life Support in Older People: An Ethical Dilemma*, 38 J. AM. GERIATRICS SOC'Y 710, 712 (1990) ("[W]hen suffering is unlikely to abate, a decision [by a depressed patient] that death is preferable to life may not necessarily be unreasonable.").

[73] *See* Model Act § 5(b).

cuss all medical treatments that might improve the patient's condition or prognosis that are practicably available, including treatment for pain, and their benefits and burdens; to offer the patient the opportunity to consult with social workers about social services that may improve his or her condition; and to advise the patient of the options for ending his or her life and their benefits and burdens. For a request to be fully informed, the patient must understand all of this information and make a reasoned decision to seek suicide. Section 3(a)(3)(B) is intended to ensure active decisionmaking by the patient; passive acquiescence in the recommendations of others would not constitute a fully informed and reasoned decision.

Section 3(a)(3)(C) requires that the patient's request be voluntary, meaning that it is made independently, free from coercion or undue influence. The patient may consider the suggestions and recommendations of others, including the responsible physician, but the patient's choice must be his or her own decision.[74]

Finally, section 3(a)(3)(D) requires that the patient's request be enduring. Ideally, the patient will have discussed physician-assisted suicide with a number of individuals on multiple occasions. At a minimum, however, the request must be stated to the responsible physician on at least two occasions that are at least two weeks apart, without self-contradiction during that interval. The two-week period is an attempt to balance the prevention of hasty decisionmaking against the prolonging of unbearable suffering.

Section 3(a) places the responsibility on the responsible physician to ensure that all of its requirements are met. In order to provide the physician with considerable advance assurance that he or she can avoid litigation attempting to second-guess his or her determinations,[75] the statute makes the physician's standard entirely subjective: the physician need have only an "honest belief" that the elements of section 3(a) have been met in the particular case. However, to compensate for the lack of any

---

[74] *Cf.* RESTATEMENT (SECOND) OF CONTRACTS § 177(1) (1981) ("Undue influence is unfair persuasion of a party who is under the domination of the person exercising the persuasion or who by virtue of the relation between them is justified in assuming that that person will not act in a manner inconsistent with his welfare."); Maurath v. Sickles, 586 S.W.2d 723, 730 (Mo. Ct. App. 1979) (holding that undue influence in probate context is influence that destroys the free choice of the person making the will).

[75] *See supra* text accompanying note 44.

requirement of reasonableness, the responsible physician enjoys the protection conferred by the statute only if he or she also satisfies the procedural requirements of sections 4, 5, and 6,[76] which are designed to produce and preserve independent corroboration that the physician's belief is not merely honest or reasonable, but accurate. If the responsible physician materially complies with these requirements[77] and there is no proof that he or she lacked the requisite honest belief, he or she is protected from liability for assisting in a suicide.[78] As discussed below,[79] however, the responsible physician and other participants are not relieved of any liability that they may otherwise incur as a result of any malpractice that they commit in the process of assisting in a suicide.

D. *Procedures to be Followed Before and After a Patient Receives Assistance in Suicide*

Section 4 outlines the information that the responsible physician must present to the patient in order to ensure that the patient's decision is fully informed and reasoned. Section 4(a) requires the responsible physician to offer the patient any medical care that may cure or palliate the illness or relieve its symptoms. Hospice care must be offered if available, but treatments that are inconsistent with accepted medical practice or impracticable need not be.[80] Section 4(b) requires the responsible physician to make a social worker available to the patient to discuss non-medical options that might change the patient's decision to seek suicide.

---

[76] *See* Model Act § 3(a).

[77] *Cf.* 2 E. ALLAN FARNSWORTH, FARNSWORTH ON CONTRACTS § 8.16 (1990) (discussing material breach and substantial performance in contract law).

[78] *See* Model Act §§ 3(a), 13(a)–(b).

[79] *See infra* part IV.H.

[80] Sometimes treatments or other services will be impracticable because the patient lacks the resources or health insurance necessary to pay for them. Such a situation presents health care providers and patients with a painful ethical dilemma. On the one hand, it seems plainly wrong for a patient to be forced to consider suicide because of a lack of ability to obtain treatments or services that might mitigate his or her condition or circumstances. On the other hand, if there is no way to right this wrong in a particular situation, it seems doubly wrong to deny the patient the medical means of suicide that he or she has requested. We believe that if society fails to meet its moral obligation to provide appropriate health care and other services to all its citizens, it cannot justifiably deny individuals relief from conditions that they find all the more unbearable because of society's moral failure.

The responsible physician must suggest to the patient under section 4(c) that he or she consult family members about the decision to request assistance in suicide, but the patient need not do so. Although mandatory family notification has been upheld against constitutional challenges in similarly sensitive situations,[81] we believe that competent, adult patients should not be required to notify family members of their intended suicide against their will. The items required to be discussed by section 4(d) have been mentioned previously,[82] but that section also requires a recorded or documented account of the discussion with two witnesses who are entitled to question the responsible physician and the patient.

Section 5 contains the corroboration requirements. Section 5(a) requires a second medical opinion as to the patient's diagnosis and prognosis, while section 5(b) requires a combination medical-factual opinion as to the patient's qualifications for physician-assisted suicide under section 3(a)(3). Broadly worded, unsupported opinions should be insufficient to enable the responsible physician to proceed; instead, each opinion should evidence a thorough investigation and demonstrate that the patient meets the statutory standards. An opinion that conflicts with the responsible physician's opinion should prevent the responsible physician from proceeding with an assisted suicide, at least until circumstances change substantially and a consultant then agrees with the responsible physician's opinion.

Finally, section 6 requires the responsible physician to document promptly the provision of medical means of suicide to a patient, both in the patient's records and with the state's regulatory authority.

E. *Presence at the Patient's Death*

Ending one's life in solitude can be a lonely and frightening undertaking, fraught with uncertainty, ambivalence, and opportunities for failure. We hope that the responsible physician will be present at the patient's death in order to reassure the patient and to make certain that the process is carried out effectively. Section 3(b) allows, but does not require, the physician to be

---

[81] *Cf.* Planned Parenthood of Southeastern Pa. v. Casey, 112 S. Ct. 2791, 2832 (1992) (upholding requirement of parental notification before minor obtains abortion).

[82] *See supra* part IV.C.

present if the patient so desires, and section 7(a) also allows the presence of any other persons selected by the patient. Each section requires only that the final physical act of administering the means of suicide be the knowing, intentional, and voluntary act of the patient.

F. *Monitoring and Enforcement*

The submission of reports by responsible physicians allows the state Department of Public Health (or a similar regulatory agency) to collect the data (specified in section 8(a)) necessary to improve the statute's operation and to make the annual public report of its effectiveness required by section 9(d). For purposes of tracking the operation of the statute, it would be desirable to determine how often and under what circumstances medical means of suicide were actually used by patients to end their lives. However, because the responsible physician need not be present at the patient's death, and because the physician who signs the death certificate may not be the same physician who provided the deceased with the means of suicide, there appears to be no way of accurately determining the extent to which medical means of suicide are actually used.

A physician's report must not include the patient's name for reasons of privacy, but section 8(b) requires a coded link between the report and the patient's name, which may be used if legal or ethical questions should arise after the patient's death.

Section 9 requires the agency to monitor and enforce the requirements of the statute and grants the agency rulemaking authority. The statute proceeds on the assumption that it is impossible in such a complex field to deal in advance with all possible problems by a legislative act. We believe that a reasonable solution is to enact the legislation and then to provide an administrative body with the power to respond to new patterns of problems through the regulatory rulemaking process.

G. *Confidentiality, Conscientious Objection, and Discrimination*

To protect the privacy and confidentiality of everyone involved in a particular physician-assisted suicide, section 10(a) declares that any information about a patient must be kept confidential.

Section 10(b) further specifies that a responsible physician's report on file with the regulatory agency is also confidential and is not subject to the customary state statutes regarding public records.

Section 11 protects the decisions of physicians, hospital employees, and hospitals themselves to refuse to participate in physician-assisted suicide on grounds of conscience. A hospital or other institution may forbid physician-assisted suicide on its premises or within its jurisdiction if the institution notifies its staff in advance of the policy.

Finally, section 12 protects patients from discrimination by physicians, institutions, and insurers. No health care provider or insurer is permitted to require any patient to request physician-assisted suicide as a condition of eligibility for services, benefits, or insurance. At the same time, section 12 protects patients from discrimination (including the voiding of life insurance policies) because they have chosen to pursue assisted suicide. Unless physicians, institutions, and insurers opt out for reasons of conscience under section 11, they must honor patients' choices to seek or avoid assistance in suicide.

H. *Liability and Sanctions*

Section 13 protects those who participate in physician-assisted suicide from the types of liabilities identified in Part I.B. of this Article. The protection of section 13(a), however, is limited to the mere fact that a person has participated in an assisted suicide; he or she may not be convicted of homicide, for example, solely on the basis that he or she provided deadly drugs to a patient who committed suicide. On the other hand, section 13(c) notes that the statute does not limit the civil or criminal liability of any person for intentional or negligent actions merely because those actions were part of a physician-assisted suicide. Thus if a responsible physician or consulting physician commits malpractice by erroneously diagnosing a patient's condition, he or she is liable for the damages caused by that malpractice. The responsible physician is not, however, stripped of protection against liability for assisting in a suicide per se unless he or she has failed to meet the requirements of one or more sections of the statute.[83]

---

[83] *See* Model Act § 13(a).

Section 14 declares that a willful violation of a provision of section 3, 4, 5, 6, or 7 is a crime (the precise grade of the crime is left to the individual state). Whether an action results in the death of a patient or not is immaterial. Of course, a violation of one of these provisions may also render a person liable under another provision of law; for example, a responsible physician who does not comply in all material respects with sections 4, 5, and 6 does not enjoy the protection from liability for assisting in a suicide that section 13 otherwise affords. In appropriate cases, section 14 provides a prosecutor with a method for enforcing the statute that falls short of a prosecution for homicide or assisting in a suicide.

As for other wrongful acts, such as coercing a person to request or use medical means of suicide, section 13(c) leaves the definition of offenses and the imposition of sanctions to existing law.

V. Conclusion

Physician-assisted suicide has become a subject of increasingly widespread and intense public and professional debate. A growing array of efforts is also underway to make physician-assisted suicide available under the law. As noted in Part III, Oregon recently adopted legislation to allow physician-assisted suicide. Constitutional challenges to laws prohibiting assisted suicide in Washington, Michigan, and New York have recently wound their way through the courts.[84] Legislation to permit physician-assisted suicide has been introduced recently in a number of state legislatures.[85] As these efforts approach fruition, it becomes increasingly important that debates about physician-assisted suicide address concrete issues of morality and policy design. Supporters of physician-assisted suicide have a special responsibility to propose specific, detailed proposals for a well-regulated and suitably circumscribed practice. We intend the statute presented below to help meet that responsibility.

---

[84] *See supra* note 24.

[85] *See, e.g.*, Cal. A.B. 1080, 1995-96 Reg. Sess.; Colo. H.B. 1308, 60th Gen. Ass., 1st Reg. Sess. (1995); Mass. H.B. 3173, 179th Gen. Ct., 1st Ann. Sess. (1995); N.H. H.B. 339, 1995 Reg. Sess.; N.Y. S.B. 1683, 218th Gen. Ass., 1st Reg. Sess. (1995); Wis. A.B. 174, 92d Leg. Sess., 1995-96 Reg. Sess.

A MODEL STATE ACT TO AUTHORIZE AND REGULATE PHYSICIAN-ASSISTED SUICIDE

SECTION 1. STATEMENT OF PURPOSE

The principal purpose of this Act is to enable an individual who requests it to receive assistance from a physician in obtaining the medical means for that individual to end his or her life when he or she suffers from a terminal illness or from a bodily illness that is intractable and unbearable. Its further purposes are (a) to ensure that the request for such assistance is complied with only when it is fully informed, reasoned, free of undue influence from any person, and not the result of a distortion of judgment due to clinical depression or any other mental illness, and (b) to establish mechanisms for continuing oversight and regulation of the process for providing such assistance. The provisions of this Act should be liberally construed to further these purposes.

SECTION 2. DEFINITIONS

As used in this Act,

(a) "Commissioner" means the Commissioner of the Department.

(b) "Department" means the Department of Public Health [or similar state agency].

(c) "Health care facility" means a hospital, hospice, nursing home, long-term residential care facility, or other institution providing medical services and licensed or operated in accordance with the law of this state or the United States.

(d) "Intractable and unbearable illness" means a bodily disorder (1) that cannot be cured or successfully palliated, and (2) that causes such severe suffering that a patient prefers death.

(e) "Medical means of suicide" means medical substances or devices that the responsible physician prescribes for or

supplies to a patient for the purpose of enabling the patient to end his or her own life. "Providing medical means of suicide" includes providing a prescription therefor.

(f) "Patient's medical record" means (1) in the case of a patient who is in a health care facility, the record of the patient's medical care that such facility is required by law or professional standards to compile and maintain, and (2) in the case of a patient who is not in such a facility, the record of the patient's medical care that the responsible physician is required by law or professional standards to compile and maintain.

(g) "Person" includes any individual, corporation, professional corporation, partnership, unincorporated association, government, government agency, or any other legal or commercial entity.

(h) "Responsible physician" means the physician, licensed to practice medicine in this state, who (1) has full or partial responsibility for treatment of a patient who is terminally ill or intractably and unbearably ill, and (2) takes responsibility for providing medical means of suicide to the patient.

(*i*) "Terminal illness" means a bodily disorder that is likely to cause a patient's death within six months.

SECTION 3. AUTHORIZATION TO PROVIDE ASSISTANCE

(a) It is lawful for a responsible physician who complies in all material respects with Sections 4, 5, and 6 of this Act to provide a patient with medical means of suicide, provided that the responsible physician acts on the basis of an honest belief that

(1) the patient is eighteen years of age or older;

(2) the patient has a terminal illness or an intractable and unbearable illness; and

(3) the patient has made a request of the responsible physician to provide medical means of suicide, which request

(A) is not the result of a distortion of the patient's judgment due to clinical depression or any other mental illness;

(B) represents the patient's reasoned choice based on an understanding of the information that the responsible physician has provided to the patient pursuant to Section 4(d) of this Act concerning the patient's medical condition and medical options;

(C) has been made free of undue influence by any person; and

(D) has been repeated without self-contradiction by the patient on two separate occasions at least fourteen days apart, the last of which is no more than seventy-two hours before the responsible physician provides the patient with the medical means of suicide.

(b) A responsible physician who has provided a patient with medical means of suicide in accordance with the provisions of this Act may, if the patient so requests, be present and assist the patient at the time that the patient makes use of such means, provided that the actual use of such means is the knowing, intentional, and voluntary physical act of the patient.

SECTION 4. DISCUSSION WITH PATIENT AND DOCUMENTATION

Before providing medical means of suicide to a patient pursuant to Section 3 of this Act, the responsible physician shall

(a) offer to the patient all medical care, including hospice care if available, that is consistent with accepted clinical practice and that can practicably be made available to the patient for the purpose of curing or palliating the patient's illness

or alleviating symptoms, including pain and other discomfort;

(b) offer the patient the opportunity to consult with a social worker or other individual trained and experienced in providing social services to determine whether services are available to the patient that could improve the patient's circumstances sufficiently to cause the patient to reconsider his or her request for medical means of suicide;

(c) counsel the patient to inform the patient's family of the request if the patient has not already done so and the responsible physician believes that doing so would be in the patient's interest; and

(d) supply to and discuss with the patient all available medical information that is necessary to provide the basis for a reasoned decision concerning a request for medical means of suicide, including all such information regarding the patient's diagnosis and prognosis, the medical treatment options and the medical means of suicide that can be made available to the patient, and their benefits and burdens, all in accordance with the following procedures:

> **(1) at least two adult individuals must witness the discussion required by this paragraph (d), at least one of whom (A) is not affiliated with any person that is involved in the care of the patient, and (B) does not stand to benefit personally in any way from the patient's death;**
>
> **(2) the responsible physician shall inform each witness that he or she may question the responsible physician and the patient to ascertain that the patient has, in fact, heard and understood all of the material information discussed pursuant to this paragraph (d); and**
>
> **(3) the responsible physician shall document the discussion with the patient held pursuant to this paragraph (d), using one of the following methods:**

(A) an audio tape or a video tape of the discussion, during which the witnesses acknowledge their presence; or

(B) a written summary of the discussion that the patient reads and signs and that the witnesses attest in writing to be accurate.

The documentation required by this subparagraph (3) must be included and retained with the patient's medical record, and access to and disclosure of such records and copies of them are governed by the provisions of Section 10 of this Act.

SECTION 5. PROFESSIONAL CONSULTATION AND DOCUMENTATION

Before providing medical means of suicide to a patient pursuant to Section 3 of this Act, the responsible physician shall

(a) secure a written opinion from a consulting physician who has examined the patient and is qualified to make such an assessment that the patient is suffering from a terminal illness or an intractable and unbearable illness;

(b) secure a written opinion from a licensed psychiatrist, clinical psychologist, or psychiatric social worker who has examined the patient and is qualified to make such an assessment that the patient has requested medical means of suicide and that the patient's request meets the criteria set forth in Sections 3(a)(3)(A), 3(a)(3)(B), and 3(a)(3)(C) of this Act to the effect that the request is not the result of a distortion of the patient's judgment due to clinical depression or any other mental illness, is reasoned, is fully informed, and is free of undue influence by any person; and

(c) place the written opinions described in paragraphs (a) and (b) of this section in the patient's medical record.

SECTION 6. RECORDING AND REPORTING BY THE RESPONSIBLE PHYSICIAN

Promptly after providing medical means of suicide to a patient, the responsible physician shall (a) record the provision of such means in the patient's medical record, (b) submit a report to the Commissioner on such form as the Commissioner may require pursuant to Section 8(a) of this Act, and (c) place a copy of such report in the patient's medical record.

SECTION 7. ACTIONS BY PERSONS OTHER THAN THE RESPONSIBLE PHYSICIAN

(a) An individual who acts on the basis of an honest belief that the requirements of this Act have been or are being met may, if the patient so requests, be present and assist at the time that the patient makes use of medical means of suicide, provided that the actual use of such means is the knowing, intentional, and voluntary physical act of the patient.

(b) A licensed pharmacist, acting in accordance with the laws and regulations of this state and the United States that govern the dispensing of prescription drugs and devices and controlled substances, may dispense medical means of suicide to a person who the pharmacist reasonably believes presents a valid prescription for such means.

(c) An individual who acts on the basis of an honest belief that the requirements of this Act have been or are being met may counsel or assist the responsible physician in providing medical means of suicide to a patient.

SECTION 8. RECORD KEEPING BY THE DEPARTMENT

(a) The Commissioner shall by regulation specify a form of report to be submitted by physicians pursuant to Section 6(b) of this Act in order to provide the Department with such data regarding the provision of medical means of suicide as the Commissioner determines to be necessary or appropriate to enable effective oversight and regulation of the operation

of this Act. Such report shall include, at a minimum, the following information:

(1) the patient's diagnosis, prognosis, and the alternative medical treatments, consistent with accepted clinical practice, that the responsible physician advised the patient were practicably available;

(2) the date on which and the name of the health care facility or other place where the responsible physician complied with the patient's request for medical means of suicide, the medical means of suicide that were prescribed or otherwise provided, and the method of recording the discussion required by Section 4(d) of this Act;

(3) the patient's vital statistics, including county of residence, age, sex, race, and marital status;

(4) the type of medical insurance and name of insurer of the patient, if any;

(5) the names of the responsible physician, the medical and mental health consultants who delivered opinions pursuant to Section 5 of this Act, and the witnesses required by Section 4(d) of this Act; and

(6) the location of the patient's medical record.

(b) The Commissioner shall require that the report described in paragraph (a) of this section not include the name of the patient but shall provide by regulation for an anonymous coding or reference system that enables the Commissioner or the responsible physician to associate such report with the patient's medical record.

SECTION 9. ENFORCEMENT AND REPORTING BY THE DEPARTMENT

(a) The Commissioner shall enforce the provisions of this Act and shall report to the Attorney General and the appro-

priate board of registration [or similar state agency] any violation of its provisions.

(b) The Commissioner shall promulgate such rules and regulations as the Commissioner determines to be necessary or appropriate to implement and achieve the purposes of this Act and shall, at least ninety days prior to adopting any rule or regulation affecting the conduct of a physician acting under the provisions of this Act, submit such proposed rule or regulation to the Board of Registration in Medicine [or similar state agency] for such Board's review and advice.

(c) The Board of Registration in Medicine [or similar state agency] may promulgate no rule or regulation inconsistent with the provisions of this Act or with the rules and regulations of the Department promulgated under it and shall, at least ninety days prior to adopting any rule or regulation affecting the conduct of a physician acting under the provisions of this Act, submit such proposed rule or regulation to the Commissioner for the Commissioner's review and advice.

(d) The Commissioner shall report to the Legislature annually concerning the operation of this Act and the achievement of its stated purposes. The report of the Commissioner shall be made available to the public upon its submission to the Legislature. In order to facilitate such annual reporting, the Commissioner may collect and review such information as the Commissioner determines to be helpful to the Department, the Board of Registration in Medicine [or similar state agency], or the Legislature and may by regulation require the submission of such information to the Department.

SECTION 10. CONFIDENTIALITY OF RECORDS AND REPORTS

(a) The information that a person acting under this Act obtains from or about a patient is confidential and may not be disclosed to any other person without the patient's consent or the consent of a person with lawful authority to act on the patient's behalf, except as this Act or any other provision of law may otherwise require.

(b) The report that a responsible physician files with the Department pursuant to Section 6(b) of this Act is confidential, is not a public record, and is not subject to the provisions of [the state public records statute or freedom of information act].

SECTION 11. PROVIDER'S FREEDOM OF CONSCIENCE

(a) No individual who is conscientiously opposed to providing a patient with medical means of suicide may be required to do so or to assist a responsible physician in doing so.

(b) A health care facility that has adopted a policy opposed to providing patients with medical means of suicide and has given reasonable notice of such policy to its staff members may prohibit such staff members from providing such means to a patient who is within its facilities or under its care.

SECTION 12. PATIENT'S FREEDOM FROM DISCRIMINATION

(a) No physician, health care facility, health care service plan, provider of health or disability insurance, self-insured employee health care benefit plan, or hospital service plan may require any individual to request medical means of suicide as a condition of eligibility for service, benefits, or insurance. No such physician or entity may refuse to provide medical services or medical benefits to an individual because such individual has requested medical means of suicide, except as Section 11 of this Act permits.

(b) A patient's use of medical means of suicide to end such patient's life in compliance with the applicable provisions of this Act shall not be considered suicide for the purpose of voiding a policy of insurance on the life of such patient.

SECTION 13. LIABILITY

(a) No person who has acted in compliance with the applicable provisions of this Act in providing medical means of

suicide to an individual shall be subject to civil or criminal liability therefor.

(b) No individual who has acted in compliance with the applicable provisions of this Act in providing medical means of suicide to a patient shall be subject therefor to professional sanction, loss of employment, or loss of privileges, provided that such action does not violate a policy of a health care facility that complies with Section 11(b) of this Act.

(c) Except as provided in paragraphs (a) and (b) of this section, this Act does not limit the civil, criminal, or disciplinary liability of any person for intentional or negligent misconduct.

SECTION 14. CRIMINAL PENALTIES

In addition to any other civil, criminal, or disciplinary liability that he or she may otherwise incur thereby, an individual who willfully violates Section 3, 4, 5, 6, or 7 of this Act is guilty of a [specify grade of offense].

ARTICLE

THE LAST ARTICLE ABOUT THE LANGUAGE OF ERISA PREEMPTION? A CASE STUDY OF THE FAILURE OF TEXTUALISM

CATHERINE L. FISK*

The Employee Retirement Income Security Act of 1974 has been held to preempt a vast array of state statutes and common law, ranging from family leave programs to health care finance reforms. In this Article, Professor Fisk argues that the Supreme Court's misguided faith in textualist methods of interpreting ERISA's preemption provisions has produced doctrinal confusion and unintended public policy. While she endorses the Court's move last Term to a more pragmatic approach to ERISA preemption, Professor Fisk's account of the development of ERISA preemption doctrine helps to explain how textualist methods of statutory interpretation may have significant—and oftentimes unintended—effects on the development of law and public policy.

The future of health care reform, if it has any, seems now to lie in the states. Yet the conventional wisdom is that the states are powerless to act unless Congress grants so-called "ERISA waivers." That is, in order for states to have authority to reform private health care payment systems, Congress must amend section 514(a) of the Employee Retirement Income Security Act of 1974 (ERISA, or "the Act"), which broadly preempts state laws that "relate to" employee benefit plans.[1] Many states enacted

---

* Associate Professor, Loyola Law School, Los Angeles. A.B., Princeton University, 1983; J.D., University of California, Berkeley, 1986. Thanks to Ellen Aprill, Peter Carstensen, Erwin Chemerinsky, and Karl Manheim for commenting on drafts of this Article; to Kathleen Baxter-Walker, Jadene Derr, Shari Faris, and Geoffrey Moore for research assistance; and to Carin Clauss, Jane Schacter, and Clyde Spillenger for helping me to clarify my thoughts at an early stage.

This is part of a study on the implementation of ERISA begun during graduate study at the University of Wisconsin. Another part was published last year. Catherine L. Fisk, *Lochner Redux: The Renaissance of Laissez-Faire Contract in the Federal Common Law of Employee Benefits*, 56 OHIO ST. L.J. 153 (1995). I am grateful for financial support from the University of Wisconsin Law School and from Loyola Law School. The author of this Article may be reached by e-mail at CFISK@LMULAW.LMU.EDU.

[1] 29 U.S.C. § 1144(a) (1988). An "ERISA waiver" is a shorthand way of referring to an amendment to the ERISA preemption provision to eliminate preemption of state law in a particular circumstance. Unlike Medicaid, where an administrative agency decides on a state's request for an exemption from federal law, there is no agency to fine-tune the relationship between state and federal law in regard to ERISA; state reforms must await action from Congress. On the general problem of ERISA waivers and state health care reform, see Devon P. Groves, *ERISA Waivers and State Health Care Reform*, 28 COLUM. J. L. & SOC. PROBS. 609 (1995); HEALTH, EDUCATION, AND

health care reform legislation, and many believed that such legislation was preempted by ERISA.[2] Although several states have sought congressional exemptions from ERISA preemption, none has been granted, except to Hawaii in 1983.[3] Thus, ERISA preemption has thwarted reform efforts in a large number of states.[4]

ERISA's sweeping preemption of state laws regulating health care payment is odd, because ERISA itself has little to do with the regulation of health finance; it simply imposes fiduciary and reporting obligations on private employee benefit plans.[5] ERISA does not require employers to provide health insurance or any

---

HUMAN SERVICES DIVISION, U.S. GAO, EMPLOYER-BASED HEALTH PLANS: ISSUES, TRENDS, AND CHALLENGES POSED BY ERISA, REP. NO. GAO/HEHS 95-167 (July 1995).

[2] *See, e.g.*, Edwin Chen, *States Take Up Health Reform Fight*, L.A. TIMES, Sept. 30, 1994, at A18; William Claiborne, *Health Reform on the Go, State by State*, WASH. POST, Nov. 26, 1993, at A29.

[3] The House Committee on Education and Labor recommended in 1993 that § 514 be amended to grant waivers to Maryland, Minnesota, New York, and Hawaii. H.R. REP. NO. 111, 103d Cong., 1st Sess. 109–12 (1993), *reprinted in* 1993 U.S.C.C.A.N. 378. However, no waivers were enacted. Omnibus Budget Reconciliation Act of 1993, Pub L. 103-66, 107 Stat. 512 (1993).

[4] Oregon and Washington both enacted ambitious health care financing reforms, but state officials reportedly believed that congressional waivers of ERISA preemption were essential to the validity of the state laws. *See* 139 CONG. REC. E3126 (Nov. 24, 1993) (statement of Rep. Wyden (D-Or.)); John Kitzhaber & Mark Gibson, *The Crisis in Health Care: The Oregon Health Plan as a Strategy for Change*, 3 STAN. L. & POL'Y REV. 64 (1991). Both Oregon and Washington sought waivers, but neither bill was enacted, and both states' reforms could not take effect without the waivers. *See id.* 139 CONG. REC. E1974-02 (daily ed. Aug. 4, 1993) (statement of Rep. Kriedler (D-Or.)); 138 CONG. REC. E3059-02 (daily ed. Oct. 5, 1992) (statement of Rep. Wyden). As one Oregon official put it: "It goes down the toilet without a waiver." Chen, *supra* note 2. Washington repealed its law recently, just two years after enacting it. *Washington Governor Signs Bills Repealing Health Care Reform Law*, L.A. TIMES, May 10, 1995, at A25. Florida, Hawaii, Minnesota, Oregon and Vermont enacted health care reform proposals that depended in part on waivers from ERISA preemption. Milt Freudenheim, *States Seek Aid for the Uninsured*, N.Y. TIMES, June 23, 1992, at D2. Massachusetts enacted similar legislation which was never enforced in part because of concerns about ERISA preemption. *See* Mary Anne Bobinski, *Unhealthy Federalism: Barriers to Increasing Health Care Access for the Uninsured*, 24 U.C. DAVIS L. REV. 255, 305–24 (1990). An initiative on the November 1994 ballot in California that would have established a state-funded, single-payer health care system was considered likely by the State Legislative Analyst to require a change in the ERISA preemption provision. ANALYSIS BY THE LEGISLATIVE ANALYST, CALIFORNIA BALLOT PAMPHLET 45 (November 1994). *See generally* Jerry Mashaw, *Taking Federalism Seriously: The Case for State-Led Health Care Reform*, DOMESTIC AFF. (Winter 1993-94), *reprinted in* 140 CONG. REC. E59,957 (daily ed. July 28, 1994) (statement of Sen. Phil Gramm (R-Tex.)); *Look Behind Today's Worst Health Insurance Horrors and the Same Monster Lurks . . . . ERISA: The Law That Ate Health Care Reform.* CAL. LAW., May 1993, at 40.

[5] In contrast to ERISA's sparse regulation of health and other benefit plans, ERISA comprehensively regulates pension plans. Thus, sweeping preemption of state laws relating to pension plans has not created the regulatory void that preemption of laws relating to nonpension (welfare benefit) plans has created.

other benefit; it does not regulate what employers can charge for benefits; it does not prevent employers from eliminating benefits (except pensions). In divesting states of authority to regulate, ERISA preemption has created an enormous, unanticipated "regulatory vacuum": ERISA has been interpreted to preempt a wide variety of state common law and statutes, including family leave[6] and workers' compensation programs,[7] prevailing wage laws,[8] provisions regulating working conditions of apprentices,[9] mechanics' liens,[10] statutes allocating damages in tort[11] and wrongful death[12] actions, taxes on hospitals,[13] novel state efforts to address the perceived crisis of the unavailability of health insurance,[14] and even certain medical malpractice claims.[15] Moreover, ERISA eliminates state claims even when ERISA itself provides

---

[6] *See* Gabrielle Lessard, Note, *Conflicting Demands Meet Conflicts of Laws: ERISA Preemption of Wisconsin's Family and Medical Leave Act*, 1992 WIS. L. REV. 809.

[7] *See* District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580 (1992); Benson v. Wyatt Cafeterias, Inc., 780 F. Supp. 1132 (N.D. Tex. 1991). *But see* Eurine v. Wyatt Cafeterias, Inc., No. 3-91-0408-H, 1991 WL 206172 (N.D. Tex., Aug. 21, 1991), *amending*, 13 Employee Benefits Cases (BNA) 2728.

[8] Dillingham Constr. N.A. Inc. v. County of Sonoma, 57 F.3d 712 (9th Cir. 1995); Chamber of Commerce v. Bragdon, 769 F. Supp. 1537 (N.D. Cal. 1991), *aff'd*, 64 F.3d 497 (9th Cir. 1995).

[9] *See infra* text accompanying notes 200–214.

[10] Trustees of the Elec. Workers Health & Welfare Trust v. Marjo Corp., 988 F.2d 865, 868 (9th Cir. 1993).

[11] Travitz v. Northeast Dept. ILGWU Health & Welfare Fund, 13 F.3d 704, 709–10 (3d Cir. 1994).

[12] McInnis v. Provident Life & Accident Ins. Co., 21 F.3d 586, 589–90 (4th Cir. 1994).

[13] NYSA-ILA Med. v. Axelrod, 27 F.3d 823 (2d Cir. 1994), *vacated*, Chassin v. NYSA-ILA Med., 115 S. Ct. 1819 (1995); Connecticut Hosp. Assn v. Pogue, 870 F. Supp. 444 (D. Conn. 1994), *rev'd*, Connecticut Hosp. Assn. v. Weltman, 66 F.3d 413 (2d Cir. 1995). Both of these cases were overturned on the strength of the Supreme Court's recent decision in N.Y.S. Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Co., 115 S. Ct. 1671 (1995) [hereinafter *New York Blues*].

[14] *See supra* note 4. In *New York Blues*, 115 S. Ct. 1671 (1995), the Supreme Court rejected a challenge to the system of differing surcharges that New York imposes on hospital rates. The surcharges depend on whether the payer is a private health insurance company, an HMO, Blue Cross/Blue Shield, etc., in order to equalize rate advantages between those payers that allow open enrollment and use community-based risk ratings (thus insuring the otherwise uninsurable), and those that do not. *See also* Bricklayers Local No. 1 Welfare Fund v. Louisiana Health Ins. Assn., 771 F. Supp. 771 (E.D. La. 1991); General Split Corp. v. Mitchell, 523 F. Supp. 427 (E.D. Wis. 1981) (Wisconsin risk-pool statute preempted).

[15] The courts are in conflict over the extent of ERISA preemption of medical malpractice claims. The problem arises because many health plans make medical decisions in evaluating whether to provide care (in the case of an HMO) or coverage (in the case of a plan). When the medical decision proves harmful, the patient ordinarily would have a malpractice claim under state law. But since ERISA preempts state claims arising out of claims for benefits under ERISA plans, medical malpractice claims are arguably preempted. ERISA provides only contract-type damages, which are obviously inadequate to remedy the harm caused by negligent medical decisions.

no remedies.[16] Every single ERISA preemption decision from the Supreme Court has involved an effort to preempt state statutes or common law created to protect employees, consumers, or the participants or beneficiaries of employee benefit funds.[17]

It is a rich irony that ERISA, which was heralded at its enactment as significant federal protective legislation,[18] has through its preemption provision been the basis for invalidating scores of progressive state laws. This Article explains that irony. I argue that the disastrous effects of ERISA preemption are the unwanted offspring of the Supreme Court's failed twenty-year love affair with variations of textualism as the dominant mode of interpreting ERISA's preemption provision. I use "textualism" in a slightly unconventional way to refer to methods of interpretation that claim to find determinate meaning in the language, history, or structure of a statute rather than acknowledge judicial responsibility for augmenting legislation to deal with unforeseen

---

Some courts have held ERISA to preempt malpractice claims. Tolton v. American Biodyne, Inc., 854 F. Supp. 505 (N.D. Ohio 1993); Corcoran v. United Healthcare, Inc., 965 F.2d 1321 (5th Cir. 1992), *cert. denied*, 113 S. Ct. 812 (1992). Both courts characterized the claim as relating to a "benefit determination." The most problematic cases are those involving HMOs, where the "plan" itself is basically made up of a hospital and its physicians. If ERISA preempts the claims against the HMO and its employees, the plaintiff has no one to sue for malpractice. Some courts have held all claims preempted. Kuhl v. Lincoln Nat. Health Plan, 999 F.2d 298 (8th Cir. 1993); Pomeroy v. Johns Hopkins Med. Servs., Inc., 868 F. Supp. 110 (D. Md. 1994); Rollo v. Maxicare of Louisiana, Inc., 695 F. Supp. 245 (E.D. La. 1988); Craft v. Northbrook Life Ins. Co., 813 F. Supp. 464 (S.D. Miss. 1993); Dukes v. United States Health Care Sys., Inc., 848 F. Supp. 39 (E.D. Pa. 1994); Rice v. Panchal, 875 F. Supp. 471 (N.D. Ill. 1994); Ricci v. Gooderman, 840 F. Supp. 316 (D.N.J. 1993). Some courts have held that ERISA does not preempt claims against HMOs based on vicarious liability for the negligence of physicians, even though ERISA preempted claims based on direct liability for the plan's negligence in the selection of doctors or in the administration of the plan. Pacificare v. Burrage, 59 F.3d 151 (10th Cir. 1995); Stroker v. Rubin, Civ. A. No. 94-5563, 1994 WL 719694 (E.D. Pa., Dec. 22, 1994); *see also* Dearmas v. Av-Med, Inc., 865 F. Supp. 816 (S.D. Fla. 1994); Haas v. Group Health Plan, Inc., 875 F. Supp. 544 (S.D. Ill. 1994); Jackson v. Roseman, 878 F. Supp. 820 (D. Md. 1995); Elsesser v. Hospital of the Philadelphia College of Osteopathic Med., 802 F. Supp. 1286 (E.D. Pa. 1992). One court held that ERISA preempted claims against an HMO, but it allowed claims against the physician to proceed. Altieri v. CIGNA Dental Health, Inc., 753 F. Supp. 61 (D. Conn. 1990). *See generally* Larry J. Pittman, *ERISA's Preemption Clause and the Health Care Industry: An Abdication of Judicial Law-Creating Authority*, 46 FLA. L. REV. 355 (1994); Michael A. Hiltzik, *Supreme Court Won't Allow Suit in Death Case Litigation*, L.A. TIMES, May 16, 1995, at D1.

[16] *See, e.g.*, Olson v. General Dynamics Corp., 951 F.2d 1123, 1128 (9th Cir. 1992); Phillips v. Amoco Oil Co. 799 F.2d 1464, 1470 (11th Cir. 1986).

[17] *See* cases cited *infra* note 104.

[18] H.R. REP. No. 533, 93d Cong., 2nd Sess., *reprinted in* 1974 USCCAN 4639–40, 4666–67, 4676–77; S. REP. No. 127, 93d Cong., 2nd Sess. 34 (1974), *reprinted in* 1974 USCCAN 4838–39; S. REP. No. 383, 93d Cong., 1st Sess. 81 (1974), *reprinted in* 1974 USCCAN 4890–91, 4898–4906.

circumstances.[19] Through sometimes extreme forms of textualist interpretation, the Court has transformed an ill-considered and hastily drafted legislative compromise into a matter of principle, asserting that Congress intended to leave regulation of employee benefits largely to the market and to private contract. But Congress neither foresaw nor intended that ERISA would effect this vast deregulation. Rather, it was a result of expansive judicial interpretation of the preemption provision and an unintended by-product of textualism as a method of statutory interpretation. Not only was the irrationality of ERISA preemption not the deliberate choice of Congress, it was not even the deliberate choice of the Supreme Court. Indeed, it was inconsistent with the Rehnquist Court's avowed preference for federalism.[20] Moreover, the Court's textualism generated uncertainty in the law which complicated the administration of employee benefit plans and states' regulatory efforts.[21] If ever there were a case study of the failures of textualism as a method of statutory interpretation, this is it. Fortunately, the Supreme Court last Term decided a case that suggests that the Justices have realized they erred and are taking a different approach, if not to statutory interpretation in general, at least to ERISA preemption of state law.[22]

Last Term, in *New York State Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Company* ("*New York Blues*"),[23] the Court abandoned its slavish devotion to literalist textualism in interpreting ERISA's broad preemption provision and instead

---

[19] I use "textualism" in a broader sense than it is ordinarily used. By "textualism," I mean not only strict "plain language," but also plain language aided by methods of statutory interpretation—often called intentionalist or purposivist—that purport to decide cases by looking at the legislative history to discern the legislature's intent about a provision or its purpose in enacting a provision. *See* CASS R. SUNSTEIN, AFTER THE RIGHTS REVOLUTION: RECONCEIVING THE REGULATORY STATE 123–30 (1990). I lump all three of these disparate modes of interpretation together into one category to distinguish them from a mode of interpretation that abandons the notion of a legislatively determined statutory meaning and instead recognizes that the courts are making choices with little legislative guidance. The contrast is drawn in Nicholas S. Zeppos, *Justice Scalia's Textualism: The "New" New Legal Process*, 12 CARDOZO L. REV. 1597 (1991). In the conventional typology, my suggested mode of interpretation is aligned with the new version of "legal process" and "pragmatist" theories of statutory interpretation. *See* sources cited *infra* note 253.

[20] E.g., Gregory v. Ashcroft, 501 U.S. 452, 457–61 (1991); New York v. United States, 505 U.S. 144 (1992). *See infra* text accompanying notes 32–38.

[21] *See* Richard J. Pierce, Jr., *The Supreme Court's New Hypertextualism: An Invitation to Cacaphony and Incoherence in the Administrative State*, 95 COLUM. L. REV. 749 (1995).

[22] N.Y.S. Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Co., 115 S. Ct. 1671 (1995).

[23] *Id.*

adopted a pragmatic approach. The Court will no longer look to the dictionary definition of the words of section 514, but instead will ask whether preemption of state law will serve the objectives of ERISA. This signals a long overdue and laudable reorientation in the Court's approach to ERISA preemption. Attention to the consequences of ERISA preemption, however, will force the Court to confront a significant policy issue that contributed to the preemption problem in the first place: namely, that Congress evaded the task of defining the appropriate spheres of state and federal regulation of employee benefits. Allocation of state and federal regulatory authority is a vexing issue in the area of health policy.[24]

I begin my argument, in Part I of this Article, with a brief review of preemption doctrine. I note that although implied preemption is said to raise concerns about undue interference with state authority, even express preemption provisions such as ERISA's raise the same issues. The problem, I argue, is that Congress cannot readily define the scope of preemption *ex ante* with sufficient specificity to relieve the courts of the obligation to accommodate state and federal law in each case. I then examine the ambiguities in the language and legislative history to support my claim that the apparent breadth of the ERISA preemption provision (which calls for ERISA to supersede "any and all" state laws "insofar as they . . . relate to" ERISA-covered employee benefit plans)[25] is *not* evidence that Congress intended to divest states of their traditional authority to regulate all terms of employment that happen to relate to employee benefit plans.

In Part II, I trace the evolution of the ERISA preemption doctrine in the Supreme Court. Although, as I show, the Court relied mainly on three variations of textualism, I also show that what seemed the obvious and unambiguous meaning to the Supreme Court seemed so at least partly because of unspoken assumptions about federalism and unregulated contract in the

---

[24] Allocation of state and federal regulatory authority was one of the major points of controversy during the recent debate over nationwide health care reform legislation. *See, e.g.*, Robert R. Rosenblatt, *Health Reform: Tangled Up in a Knot of Deal-Killers*, L.A. TIMES, Aug. 21, 1994, at A1 (characterizing as a "deal-killer" any proposal that would allow states to regulate employee benefits); *See generally*, Candice Hoke, *Constitutional Impediments to National Health Reform: Tenth Amendment and Spending Clause Hurdles*, 21 HAST. CONST. L.Q. 489, 499–503 (1994); Fernando R. Laguarda, Note, *Federalism Myth: States as Laboratories of Health Care Reform*, 82 GEO. L.J. 159 (1993).

[25] 29 U.S.C. § 1144(a) (1988).

law of labor relations. Put another way, the meaning that the Court gave to the preemption clause did not inhere in the text, but was put there by the Court because it believed that private, contractual approaches to problems of employee benefits were to be preferred and that the contractual scheme implicit in ERISA must be protected against undue state encroachment. By failing to recognize that the problem of ERISA preemption is one of regulatory federalism—and not one of giving life to an unambiguous statutory structure—the Justices obscured, even from themselves, the nature of the choices they were making.

Congress' decision to preempt state law without either creating a federal regulatory structure to fill the gap or instructing courts whether or how to create a federal common law to do so, put courts in the position of resolving elemental disputes about employment policy. The Supreme Court resorted to textualism in an effort to avoid explicitly making the choices that Congress had failed to make in drafting the legislation. At the close of Part II, I suggest that the problem of ERISA preemption is a consequence of the disintegration of the post-war paradigm of labor law.[26] In the New Deal-era vision that animated the liberal labor-business coalition that enacted ERISA, national legislation regulating employment was to be preferred to the inconsistent and inadequate protections of state law and the hostility of the state judiciary. But broad preemption under ERISA, as under the federal labor law, became problematic when the deficiencies of the federal law protections were revealed.[27]

---

[26] *See generally* Katherine Van Wezel Stone, *The Post-War Paradigm in American Labor Law*, 90 YALE L.J. 1509 (1981); JAMES B. ATLESON, VALUES AND ASSUMPTIONS IN AMERICAN LABOR LAW (1983).

[27] The literature criticizing ERISA preemption is large and growing. Some of the more recent and more notable contributions are Jay Conison, *ERISA and the Language of Preemption*, 72 WASH. UNIV. L. Q. 619 (1994); Paul O'Neil, *Protecting ERISA Health Care Claimants: Practical Assessment of a Neglected Issue in Health Care Reform*, 55 OHIO ST. L.J. 724 (1994); Bobinski, *supra* note 4; Leon E. Irish & Harrison C. Schaffer, *ERISA Preemption: Judicial Flexibility and Statutory Rigidity*, 19 U. MICH. J. L. REF. 109 (1985); William J. Kilberg & Paul D. Inman, *Preemption of State Laws Relating to Employee Benefit Plans: An Analysis of ERISA Section 514*, 62 TEX. L. REV. 1313 (1984). The chapter on preemption in the leading ERISA casebook is excellent, JOHN H. LANGBEIN & BRUCE A. WOLK, PENSION & EMPLOYEE BENEFIT LAW ch. 9 (2d ed. 1995), as is the summary of preemption in STEPHEN R. BRUCE, PENSION CLAIMS: RIGHTS AND OBLIGATIONS (2d ed. 1993).

There is also a literature criticizing broad preemption in labor law. *See, e.g.*, Michael H. Gottesman, *Rethinking Labor Law Preemption: State Laws Facilitating Unionization*, 7 YALE J. ON REG. 355 (1990); Eileen Silverstein, *Against Preemption in Labor Law*, 24 CONN. L. REV. 1 (1991); Lee Modjeska, *Federalism in Labor Relations—The Last Decade*, 50 OHIO ST. L. J. 487 (1989); William B. Gould IV, *When State and Federal Laws Collide: Preemption—Nightmare or Opportunity?*, 9 INDUS. REL. L.J. 4,

In Part III, I argue for a theory of statutory interpretation that would allow courts to adopt an approach to preemption that facilitates consideration of the policy consequences of broad preemption. The Supreme Court had it right when it indicated in *New York Blues* that henceforward courts should decide whether ERISA preempts state law not by asking whether the language requires it or whether Congress intended it, but by asking whether preemption makes sense as a matter of ERISA policy. To know whether preemption makes sense in light of the purpose and function of ERISA, however, courts must do what Congress failed to do: develop a preemption doctrine sensitive to the different degrees of substantive regulation that ERISA imposes on pensions as opposed to nonpension benefits.

Unlike others, I do not believe that legislative revision of the ERISA preemption provision is necessary to reorient preemption doctrine.[28] Even if a significant revision were to pass Congress (which, as I explain in Part III, is unlikely to happen), Congress would face difficult line-drawing problems that could not be resolved *ex ante*. As a practical matter, this means that the courts, rather than Congress, will have the leading role in defining the scope of ERISA preemption.

Whatever the future of health care reform at the federal level, these problems will have to be addressed under ERISA. If, as appears likely, health care reform in the 104th Congress amounts to nothing or to only slight modification of the rules on portability of benefits and preexisting condition exclusions,[29] the preemption problems will remain for health benefits as for other ERISA-covered benefits. Even if Congress were to enact more dramatic health care reform, the preemption problems will remain for child care, vacation, sick leave, apprenticeship pro-

---

19–29 (1987); Michael Shultz & John Husband, *Federal Preemption Under the NLRA: A Rule in Search of a Reason*, 62 DENV. U. L. REV. 531 (1985); Archibald Cox, *Recent Developments in Federal Labor Law Preemption*, 41 OHIO STATE L.J. 277 (1980).

[28] E.g., James D. Hutchinson & David M. Ifshin, *Federal Preemption of State Law Under the Employee Retirement Income Security Act of 1974*, 46 U. CHI. L. REV. 23, 24 (1978) (arguing that "continued whittling away of the preemptive reach of ERISA seriously threatens the regulatory scheme devised by Congress, and that it is up to Congress, not the courts, to narrow ERISA's preemption of state law where particular policy reasons make such action appropriate").

[29] For example, H.R. 995, 104th Cong., 1st Sess. (1995), would amend ERISA to provide portability of health insurance. S. 308, 104th Cong., 1st Sess. (1995) would waive ERISA preemption of certain state health reform programs in limited circumstances. There is no legislation currently pending that would exempt from ERISA preemption all state health care reform legislation.

grams, and other benefits that are administered by ERISA-covered plans.

I. The Ambiguous Language and History of the ERISA Preemption Provision

A. *Express and Implied Preemption: A Distinction Without Much Difference*

All federal statutes raise an issue of preemption of state law. The general principle of federal preemption is that, subject only to the substantive limitations on Congress's power, Congress may preempt state law to whatever extent Congress may choose.[30] For this reason, the judicial preemption inquiry is conventionally described as being a matter of discerning Congress's intent.[31] The Supreme Court has insisted that congressional intent to preempt state law be "clear" so as not to impinge unduly upon state power.[32] Thus, the Court often says that it assumes Congress does not intend federal law to supersede "the historic police powers of the States . . . unless that was the clear and manifest purpose of Congress."[33] However, Congress often does not attempt to expressly articulate its intent regarding preemption. In such circumstances, courts may infer preemptive intent either from the fact that the federal statute "occupies the field" or from the fact that state law directly conflicts with or somehow "stands as an obstacle to" the objectives of Congress.[34] Judges

---

[30] Although most cases assert that this congressional authority derives from the Supremacy Clause, recent scholarship has suggested that preemption need not always be a matter of the Supremacy Clause, but rather is derived from congressional power to enact substantive legislation. *See* Stephen A. Gardbaum, *The Nature of Preemption*, 79 Cornell. L. Rev. 767 (1994); *see also* S. Candice Hoke, *Transcending Conventional Supremacy: A Reconstruction of the Supremacy Clause*, 24 Conn. L. Rev. 829 (1992); S. Candice Hoke, *Preemption Pathologies and Civic Republican Values*, 71 B.U. L. Rev. 685 (1991).

[31] E.g., Pilot Life Ins. Co. v. Dedeaux, 481 U.S. 41, 45 (1987).

[32] *E.g.*, Rice v. Santa Fe Elevator Corp., 331 U.S. 218, 230 (1947). *Cf.* Gregory v. Ashcroft, 501 U.S. 452, 457–61 (1991) (requiring a "plain statement" by Congress).

[33] Hillsborough County v. Automated Med. Labs., Inc., 471 U.S. 707, 715 (1985); Jones v. Rath Packing Co., 430 U.S. 519, 525 (1977); Rice v. Santa Fe Elevator Corp., 331 U.S. 218, 230 (1947). The reason for this assumption is unclear, but presumably it comes from some generalized constitutional notion of the value of federalism. As I suggest in part III, I do not believe the assumption is helpful, much less compelled, either as a matter of constitutional law or sensible policy. *See infra* text accompanying notes 254–257.

[34] *E.g.*, Pacific Gas & Elec. Co. v. State Energy Resources Conservation & Dev. Comm., 461 U.S. 190, 204 (1983). Although this is the conventional "implied"

complain about these implied preemption analyses, for it is difficult to discern when Congress has occupied a field and what the scope of that field is, or when state law is an obstacle to some congressional goal.[35] Although the Supreme Court recently has insisted that Congress must make a "plain statement" when it enacts legislation that alters the balance of power between the state and federal governments,[36] it has been suggested that this "plain statement" rule may conflict with some implied preemption cases.[37] However, the conflict is nothing new; implied preemption doctrines have always been in tension with the Court's claim that congressional intent to preempt be "plain."[38]

Given the difficulties implied preemption analyses pose for judges, ERISA's express preemption provision was greeted with a sigh of relief by some commentators and judges, including initially the Supreme Court, which treated ERISA preemption as

---

preemption rule, *see generally* KENNETH STARR, ET AL., AMERICAN BAR ASSOCIATION, THE LAW OF PREEMPTION: A REPORT OF THE APPELLATE JUDGES CONFERENCE 19–30 (1991), it has been criticized. *See, e.g.*, Stephen A. Gardbaum, *The Nature of Preemption*, 79 CORNELL. L. REV. 767, 807–12 (1994); Paul Wolfson, *Preemption and Federalism: The Missing Link*, 16 HASTINGS CONST. L.Q. 69 (1988).

[35] As federal judges who authored an ABA monograph on preemption point out:

> Under occupying the field analysis, a broad legislative scheme is deemed to inform the courts that Congress, reflecting upon the interests of the states, intended all state laws touching the area to be superseded. But it may equally be reasonable to assume just the opposite—that the intrusion on state prerogatives sanctioned by the comprehensive federal program represents all that Congress considered appropriate. Further, even if one accepts the inference that courts often draw from federal statutory complexity, it remains difficult to claim that this implied intent can be taken to preempt state laws that supplement or are otherwise in harmony with the federal scheme.

Starr, *supra* note 34, at 34–35.

A similar problem exists with regard to the second form of implied preemption, the "obstacle" doctrine:

> It is unclear when, if ever, Congress has *not* balanced and compromised in enacting legislation. If every state law affecting one of the many interests reconciled by a particular federal statute were preempted under a delicate balance theory, there would seem to be little if any room for state regulatory authority. In short, lack of standards to guide this inquiry can transform a delicate balance into federal occupation of a field.

Id.

[36] Gregory v. Ashcroft, 501 U.S. 452, 470 (1991).

[37] Henry H. Drummonds, *The Sister Sovereign States: Preemption and the Second Twentieth Century Revolution in the Law of the American Workplace*, 62 FORDHAM L. REV. 469, 528 (1993).

[38] If *Gregory*'s "plain statement" rule were applied to preemption (which the Supreme Court apparently has not considered), entire bodies of preemption doctrine might be called into question. For example, since Congress did not clearly state an intent to preempt all state law regulating labor relations, many of the cases following San Diego Building Trades Council v. Garmon, 359 U.S. 236 (1959), and Lodge 76, International Association of Machinists v. Wisconsin Employment Relations Commission, 427 U.S. 132 (1976), might be of doubtful validity. *See infra* notes 237–241.

if it were straightforward.[39] Unlike many federal statutes, such as the National Labor Relations Act, ERISA expressly addresses the problem: the Act "supersede[s] any and all State laws insofar as they may now or hereafter relate to any employee benefit plan . . . ."[40] ERISA appeared to represent a clear statement of Congress's intent to preempt *all* state law. The experience of courts with ERISA preemption, however, demonstrates that interpretation of an express preemption provision can raise the same problems of regulatory federalism that implied preemption raises, at least when the federal statute applies as broadly as does ERISA.

Interpreting an express preemption provision like ERISA's, I contend, does not differ dramatically from the task of interpreting the preemptive effect of a statute without an express preemption provision. In both cases, the courts must engage in a pragmatic process of determining when the enforcement of state law is consistent with the objectives of federal regulation. Certainly in the case of ERISA, and perhaps in other areas as well, the judicial preemption analysis is less constrained by legislative direction than the "congressional intent" rhetoric would suggest. Although express preemption is the reform proposed by some scholars and judges who believe that the inferring of an intent to oust all state regulation when a federal statute seems to "occupy the field" is of questionable validity in a federal system,[41] the experience of federal courts with the ERISA preemption clause may suggest that clear statements are easier to ask for than to give or to receive. ERISA is thus evidence that the

---

[39] As the Supreme Court said in its first ERISA preemption opinion, "we are assisted by an explicit congressional statement about the pre-emptive effect of" the statute which "demonstrates that Congress . . . meant to establish pension plan regulation as exclusively a federal concern." Alessi v. Raybestos-Manhattan, Inc., 451 U.S. 504, 522–23 (1981).

[40] 29 U.S.C. § 1144(a) (1988). Although there are a number of express limits on the scope of preemption, they do little to give meaning to the main provision. First, only state laws affecting "employee benefit plans" covered by ERISA are preempted. While the term "plan" is not defined in the statute, the definition of "employee benefit plan" excludes plans maintained "solely" to comply with state workers' compensation or disability benefit laws; thus, laws relating to such plans are not preempted. § 4(b)(3), 29 U.S.C. § 1003(b)(3) (1988). Second, there are several categories of state laws that, although they relate to plans covered by ERISA, are nevertheless expressly saved from preemption; state insurance laws are the most significant among these. § 514(b)(2)(A), 29 U.S.C. § 1144(b)(2)(A) (1988). But, to limit the scope of the so-called "insurance savings" provision, ERISA states that no employee benefit plan may be deemed to be insurance for the purpose of state insurance law. § 514(b)(2)(B), 29 U.S.C. § 1144(b)(2)(B) (1988).

[41] *E.g.*, Jose L. Fernandez, *Dynamic Statutory Interpretation: Occupational Safety and Health Act Preemption and State Environmental Regulation*, 22 FLA. ST. U. L. REV. 75, 109 (1994); Starr, *supra* note 34, at 40–56; Wolfson, *supra* note 34, at 112–14.

Court's supposed new effort "to merge federalism instincts with the plain meaning doctrine of statutory interpretation" is not likely to bring great clarity to the area of preemption.[42]

B. *The Ambiguities of Section 514 and Why They Matter*

Section 514 of ERISA is fundamentally ambiguous in important respects. The Supreme Court has noted that it is "not a model of legislative drafting."[43] Nevertheless, the Court has persisted in trying to decide cases solely by reference to "the ordinary meaning of 'relate to.'"[44] The Court's emphasis on the language of section 514 invites scrutiny of the section's ambiguous meaning. How ambiguities about its effect and scope are resolved has significant consequences for state labor, insurance, health care, and consumer welfare law and policy.

Section 514(a) states that ERISA "supersede[s] any and all State laws insofar as they may now or hereafter relate to any employee benefit plan" that is not exempt from ERISA.[45] ERISA exempts from its coverage plans maintained "solely" to comply with state workers' compensation, disability, or unemployment laws, as well as government and church-sponsored plans.[46] But, apart from these exemptions and a few others not pertinent here, ERISA covers any employer or employer-union "plan, fund, or program" that provides pensions or benefits for health care, child care, vacations, sickness, disability, death, apprenticeship, training, or scholarships.[47] Thus, any law that "relates to" one of those plans is "superseded," unless it is saved by one of the savings provisions in section 514(b). Section 514(b) saves from preemption generally applicable state criminal law,[48] state law "which regulates" insurance, banking, or securities,[49] the State of Hawaii's Prepaid Health Care Act,[50] state laws regulating

---

[42] Frank L. Easterbrook, *Constitutional Law Conference*, 61 U.S.L.W. 2237, 2248 (Oct. 27, 1992). I am not certain that the Court's demand for clear statements of preemption, *see* Drummonds, *The Sister Sovereign States*, *supra* note 37, at 529; Wolfson, *Preemption and Federalism*, *supra* note 34, at 112–14, will achieve any less indeterminacy in the law than currently exists.

[43] Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724, 739 (1985).

[44] District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580 (1992).

[45] 29 U.S.C. § 1144(a) (1988).

[46] 29 U.S.C. § 1003(b) (1988).

[47] 29 U.S.C. § 1002(1), (2) (1988).

[48] 29 U.S.C. § 1144(b)(4) (1988).

[49] 29 U.S.C. § 1144(b)(2)(A) (1988).

[50] 29 U.S.C. § 1144(b)(5) (1988); HAW. REV. STAT. §§ 393-1 to 393-51 (1993).

certain entities known as Multiple Employer Welfare Arrangements,[51] and state family law orders that satisfy ERISA's definition of Qualified Domestic Relations Orders.[52]

The most obvious ambiguity concerns the meaning of "relates to." As the Supreme Court finally recognized last Term, "relates to" is a term that requires a modifier in order to have a concrete meaning,[53] and the wide spectrum of possible modifiers—directly, slightly, remotely—suggests a wide spectrum of possible meanings. Consider seven possibilities, drawn from actual or threatened ERISA preemption litigation dealing with health benefits. In each of these areas, ERISA is silent on the issues covered by the allegedly preempted state laws:

(1) A state law could "relate to" an employee benefit plan in a very direct sense, such as Hawaii's law that requires an employer to offer specified health benefits to all its employees.[54] (2) A law could relate to a plan in a less direct sense, such as provisions in the District of Columbia's workers' compensation law and Wisconsin's family and medical leave law that require employers who offer health benefits to their employees to continue those benefits while an employee is receiving workers' compensation benefits or is taking a leave to care for a family member.[55] (3) A law could relate in a still less direct sense, such as a Massachusetts law providing that every employer must pay a payroll tax to fund a state system of health benefits but exempting employers who maintain benefit plans.[56] (4) A law might relate in an even less direct sense by providing that an employer

---

[51] 29 U.S.C. § 1144(b)(6) (1988).

[52] 29 U.S.C. § 1144(b)(7) (1988). A QDRO is defined in 29 U.S.C. § 1056(d)(3)(B)(i) (1988).

[53] N.Y.S. Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Co., 115 S. Ct. 1671, 1677 (1995).

[54] Standard Oil v. Agsalud, 633 F.2d 760 (9th Cir. 1980), *aff'd*, 454 U.S. 801 (1981), held that ERISA preempts such a law. Congress later responded to this decision by exempting Hawaii's law from ERISA preemption. 29 U.S.C. § 1144(b)(5) (1988); HAW. REV. STAT. §§ 393-1 to 393-51 (1993).

[55] District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580, 583 (1992), held that ERISA preempts such a workers' compensation law. The Wisconsin Family Leave Act, which provides that health benefits must continue while an employee is on leave, raised the problem of ERISA preemption, *see* Lessard, *supra* note 6, at 834–40, but it was not litigated to a published disposition. The latter issue has been partially mooted by § 104(c) of the federal Family Medical Leave Act, which requires an employer to maintain group health benefits coverage for an employee who takes family or medical leave under the Act. 29 U.S.C. § 2601 (1988). A state law that grants more generous leave than available under the FMLA cannot require the employer to continue benefits, so an employee who opts to take the leave under the state law would not receive continued health benefits under the state law.

[56] This is what Massachusetts's so-called Pay or Play scheme would have done. *See*

must pay such a payroll tax without providing, as California's Proposition 186 would have, an exemption (which doubtless would have an impact by discouraging employers from offering private plans).[57] (5) A law, such as New York's, might impose taxes on hospitals or surcharges on hospital rates that differ depending on whether the payer is a commercial insurer, an HMO, a self-insurer, the Medicare or Medicaid systems, or Blue Cross/Blue Shield.[58] (6) A law might, as in the common law of most states, calculate damages for wrongful termination by including in the calculation of lost wages the cash value of health benefits.[59] (7) A law might simply provide, as does Michigan's business tax, that every business must pay taxes on the value it adds to the goods and services it produces; to the extent that the cost of labor, including employee benefits, is the measure of the value added, the tax would be computed by reference to the cost of providing benefits.[60]

Other examples can be found in the full range of benefits that can be offered in an ERISA plan.[61] Do state laws regulating the

---

MASS. GEN. LAWS ANN., Ch. 118F (West Supp. 1990); *see generally* Bobinski, *supra* note 4, at 305–13 & n.193.

[57] This is what the California Health Security Act (CHSA) would have done. The CHSA appeared on the November 1994 California ballot as Proposition 186. It was defeated. Dan Morain and Virginia Ellis, *California Elections/ PROPOSITIONS Voters Approve 'Three Strikes' Law, Reject Smoking Measure Proposal for Government-Run Health Care System, Gasoline Tax to Fund Rail Projects are also Defeated*, L.A. TIMES, Nov. 9, 1994, at A1. The state Legislative Analyst opined that ERISA would preempt the payroll tax aspect of the proposed legislation. ANALYSIS BY THE LEGISLATIVE ANALYST, CALIFORNIA BALLOT PAMPHLET 45 (November 1994).

[58] NYSA-ILA Medical & Clinical Servs. Fund v. Axelrod, 27 F.3d 823 (2d Cir. 1994) (ERISA preempts state tax on gross receipts of medical centers where centers are operated by an ERISA plan); N.Y.S. Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Co., 115 S. Ct. 1671, 1673–75 (1995).

[59] *See* District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580, 585 (1992) (Stevens, J., dissenting).

[60] In Thiokol Corp. v. Roberts, 858 F. Supp. 674 (W.D. Mich. 1994), the court rejected an argument that ERISA preempted the Michigan Single Business Tax, MICH. COMP. LAWS § 208.1–.145 (1995), on the ground that even though the state law referred to ERISA-covered employee benefits, it was not sufficiently related to a plan to compel preemption. In Boyle v. Anderson, 849 F. Supp. 1307 (D. Minn. 1994), the court held that ERISA did not preempt a state tax imposed on health care providers. *See generally* Kevin Matz, *ERISA's Preemption of State Tax Laws*, 61 FORDHAM L. REV. 401 (1992).

[61] To consider examples from the pension area, a state law might relate to a pension plan by prohibiting reduction of pension benefits to offset the value of other benefits, such as workers' compensation. *See* Alessi v. Raybestos-Manhattan, Inc., 451 U.S. 504 (1981) (holding that ERISA preempts such a statute). Or, a law might relate to a pension plan by according the non-employee spouse a partial share of the employee's pension benefits upon dissolution of the marriage. In General Motors Corp. v. Townsend, 468 F. Supp. 466 (E.D. Mich. 1976), the court held that a state family support order was preempted by ERISA. *But see* Stone v. Stone, 450 F. Supp. 919 (N.D. Cal. 1978), *aff'd*, 632 F.2d 740 (9th Cir. 1980), *cert. denied*, 453 U.S. 922 (1981); Cartledge

wages to be paid apprentices or the ratio of journeymen to apprentices on construction projects "relate to" plans if the apprenticeship program in question is an ERISA plan?[62] What if the operation of state law effectively dictates the amount of the employer contribution to a union-employer administered apprenticeship fund? If an employer established a child care center as an ERISA plan, would ERISA preempt state regulation of the center?[63]

ERISA simply does not unambiguously indicate whether any of the above laws bear such a relationship to a plan that they ought to be "superseded." The language of section 514(a) is, as the Supreme Court finally admitted, "unhelpful."[64] Indeed, if the state laws were unenforceable, one wonders what they would be "superseded" by, since ERISA itself says absolutely nothing about most of the subjects of the laws. Presumably, the laws would be superseded by silence, that is, by the absence of regulation. That makes little sense as a matter of statutory construction and even less sense as a matter of policy. But until last Term in *New York Blues*,[65] that was the interpretation the Court appeared to have chosen.

To decide which state laws survive preemption, a court must make interpretive choices, and those choices will profoundly affect a variety of important social and economic policy issues. Nevertheless, only three years ago, the Court took the view that

---

v. Miller, 457 F. Supp. 1146, 1158 (S.D.N.Y. 1978) (refusing to apply a "literal-minded reading of ERISA" that would thwart state family law order attaching delinquent husband's pension to pay support to wife and children). Congress fixed this problem by amending § 514(b) to exempt from preemption certain state family law orders. 29 U.S.C. § 1144(b)(7) (1988). However, the fix is only partial. Only certain state law orders are saved from preemption, and the Ninth Circuit recently held that state community property laws that give a spouse a one-half interest in the earnings of the other spouse are preempted as applied to an ERISA pension plan. Ablamis v. Roper, 937 F.2d 1450 (9th Cir. 1991). Legislation to provide for division of pension benefits upon divorce is pending in Congress. H.R. 1048, 104th Cong., 1st Sess. (1995). A law might also relate to a plan by allowing creditors to attach all of the assets of an employee, including pension benefits. *See* Mackey v. Lanier Collection Agency & Serv., Inc., 486 U.S. 825 (1988) (ERISA does not preempt such a statute).

[62] ERISA has been held to preempt both sorts of laws. *See, e.g.*, Boise Cascade Corp. v. Peterson, 939 F.2d 632 (8th Cir. 1991), *cert. denied*, 505 U.S. 1213 (1992) (ratios); Local Union 598 v. J.A. Jones Construction Co., 846 F.2d 1213 (9th Cir. 1988), *aff'd*, 488 U.S. 881 (1988) (wages). *See infra* text accompanying notes 200–215.

[63] These examples suggest that § 514 is ambiguous in the sense used by Professor Eskridge: The application of the language would lead to results that seem ridiculous or seem to contradict the historical basis for the statute. William N. Eskridge, Jr., *Dynamic Statutory Interpretation*, 135 U. PA. L. REV. 1479, 1483 (1987).

[64] N.Y.S. Conference of Blue Cross/Blue Shield Plans v. Travelers Insurance Co., 115 S. Ct. 1671, 1677 (1995).

[65] 115 S. Ct. 1671 (1995).

the plain meaning of the "relates to" language could decide cases and insisted that the dictionary definition of "relates to" ("To stand in some relation; to have bearing or concern; to pertain; to refer; to bring into association with or connection with")[66] was determinative. This approach invited the argument that any state law is preempted, since an employee benefit plan "stands in some relation" to almost any state law imaginable. State tort law, for example, which would require plan administrators to refrain from using physical force when disputing claims for benefits, might be preempted. Of course, such a construction of the statute is absurd—akin to suggesting that the request to bring every ashtray in the room means to tear them off walls, to seize them from the grip of those who are using them, and perhaps even to bring every receptacle that ever was or could be used as an ashtray.[67] The Court implicitly recognized this absurdity when it saved from preemption those laws that have only "tenuous, remote, or peripheral" connections to plans, such as many laws of "general applicability."[68]

The textualist deals with the problem of absurd results from literal readings by choosing an alternative meaning that "does least violence to the text."[69] But there is no single "alternative" meaning of "relates to" that allows a plain meaning textualist to pretend that the meaning of the statute is clear. Once the Court created the "tenuous, remote, or peripheral" exception to preemption, it in effect conceded that the "relates to" language was not itself determinative. Under the statute thus supplemented, only laws that were, in the Court's view, not *too* peripherally or *too* remotely related to plans were preempted.[70] But the diction-

---

[66] District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580, 583 (1992), quoting BLACK'S LAW DICTIONARY 1288 (6th ed. 1990).

[67] This is the example that Judge Richard Posner used in his book, THE PROBLEMS OF JURISPRUDENCE 268 (1990). *See also* Daniel A. Farber, *The Inevitability of Practical Reason: Statutes, Formalism, and Rule of Law*, 45 VAND. L. REV. 533, 544–50 (1992). On the difficulty of knowing when a particular result is "absurd," see Veronica M. Dougherty, *Absurdity and the Limits of Literalism: Defining the Absurd Result Principle in Statutory Interpretation*, 44 AM. U. L. REV. 127 (1994).

[68] *Greater Washington*, 113 S. Ct. at 583 n.1 (1992) (quoting Shaw v. Delta Air Lines, Inc., 463 U.S. 85, 100 n.21 (1983)). *See, e.g.*, Mackey v. Lanier Collection Agency & Serv., Inc., 486 U.S. 825 (1988) (generally applicable garnishment law not preempted); Aetna Life Ins. Co. v. Borges, 869 F.2d 142 (2d Cir. 1989) (escheat law not preempted though applied to unclaimed ERISA benefits), *cert. denied*, 493 U.S. 811 (1989).

[69] Green v. Bock Laundry Machine Co., 490 U.S. 504, 529 (1989) (Scalia, J., concurring).

[70] For this reason, the Court rested the holding of a case on this language only once, in *Mackey*, 486 U.S. 825 (1988). The Court's reluctance to use the language deprived

ary is then no help; only the exercise of common sense is. This approach then ceases to be "textualist" in the pure sense, and the judge instead must embark upon the task of designing a sensible preemption policy, which is precisely what the textualist thinks he is trying to avoid.

The ERISA preemption provision is ambiguous in parts other than the "relates to" language. For instance, ERISA saves from preemption state laws "which regulate insurance."[71] What is a law that "regulates insurance"? Is the tort of bad faith insurance practices such a law? In most states, it is a tort that can be committed only by an insurance company, and the law therefore "regulates insurance" in the sense that it has—or is supposed to have—an impact on the way that insurers handle claims.[72] In *Pilot Life Ins. Co. v. Dedeaux*,[73] the Supreme Court held that a cause of action based on the Mississippi common law of bad faith insurance practices was not a law which "regulates insurance" because it was not, in the Court's view, "'integral' to the insurer-insured relationship."[74] Since bringing such a suit against an insurer when it is acting as a claims processor for an employee benefit plan "related to" the plan, the Court concluded that the state law was preempted.[75] As a consequence, insurance companies face no state tort liability when handling claims through employee benefit plans.[76] This result made sense to the Court because ERISA creates other claims (none involving punitive or compensatory damages, however)[77] for the denial of claims for benefits, and the Court thought that ERISA's remedial scheme was comprehensive and ought therefore to be exclusive.[78] This

---

it of much force, since lower courts declined to rely on it either. *See, e.g.*, NYSA-ILA Medical & Clinical Servs. Fund v. Axelrod, 27 F.3d 823, 827 (2d Cir. 1994).

[71] 29 U.S.C. § 1144 (1988).

[72] Mark Gergen, *A Cautionary Tale About Contractual Good Faith in Texas*, 72 TEX. L. REV. 1235, 1250 (1994).

[73] 481 U.S. 41 (1987).

[74] *Id.* at 51 (quoting Union Labor Life Ins. Co. v. Pireno, 458 U.S. 119, 129 (1988)).

[75] *Id.* at 47–48.

[76] For criticism of *Pilot Life*, *see* Paul O'Neil, *Protecting ERISA Health Care Claimants: Practical Assessment of a Neglected Issue in Health Care Reform*, 55 OHIO ST. L.J. 723, 728–38, 763–79 (1994); Robert L. Aldisert, Note, *Blind Faith Conquers Bad Faith: Only Congress Can Save Us After* Pilot Insurance Co. v. Dedeaux, 21 LOY. L.A. L. REV. 1343 (1988); Karen L. Peterson, Comment, *ERISA Preemption of California Tort and Bad Faith Law: What's Left?*, 22 U.S.F. L. REV. 519 (1988).

[77] *See generally* George Lee Flint, Jr., *ERISA: Extracontractual Damages Mandated for Benefit Claims Actions*, 36 ARIZ. L. REV. 611 (1994) (discussing the prevailing views on the limits on extracontractual damages and arguing that ERISA does indeed authorize the award of such damages in some circumstances).

[78] *Pilot Life*, 481 U.S. at 52–56.

may or may not be a desirable result as a matter of policy, but it is not the only one compelled by the language of the statute or by its legislative history.

C. *The Origins of ERISA Preemption and the Problem of Congressional Intent*

Congress did not think very carefully about preemption when it drafted ERISA. Therefore, the ordinarily problematic process of either ascription or historical reconstruction of legislative intent and statutory purpose in section 514 is even more complicated. A legal doctrine that turns on legislative intent, as preemption does, assumes that there *is* a legislative intent to be found or that, even if an actual historic intent cannot be found, one can be imputed without undue difficulty. That is simply not true with ERISA.[79]

The actual historical evidence of congressional intent regarding preemption is sparse. Although the legislative history of ERISA is voluminous,[80] it reveals that Congress gave little thought to preemption. Careful scholarship on the history of section 514 has shown that the exceptionally broad language "was not a deeply considered result of the years of planning, negotiating, and drafting" that Congress put into ERISA.[81] As the Supreme Court has recognized, the versions of ERISA that worked their way through most of the legislative process tied the scope of preemption to the scope of ERISA regulation.[82] The House bill would have preempted state laws that "relate to the reporting and disclosure responsibilities and fiduciary responsibilities of per-

---

[79] I do not refer here to the fact that collectivities such as legislatures do not have an "intent" in the ordinary sense of the term. *See* Farber, *supra* note 67, at 551 (1992); Max Radin, *Statutory Interpretation*, 43 HARV. L. REV. 863, 870 (1930). Rather, I refer to aspects of ERISA's legislative history that make identification of a collective legislative "intent" especially problematic.

[80] The legislative history of ERISA up to 1974 has been compiled and published in a three-volume set. SUBCOMM. ON LABOR OF THE SENATE COMM. ON LABOR AND PUB. WELFARE, 94TH CONG., 2D SESS., LEGISLATIVE HISTORY OF THE EMPLOYEE RETIREMENT INCOME SECURITY ACT OF 1974 (Comm. Print 1976) ("ERISA Legislative History").

[81] Daniel C. Schaffer & Daniel M. Fox, *Semi-Preemption in ERISA: Legislative Process and Health Policy*, 7 AM. J. TAX POL'Y 47, 48 (1988). Much of the discussion of legislative history that follows is drawn from this excellent article, and from another superb study, Leon E. Irish & Harrison J. Cohen, *ERISA Preemption: Judicial Flexibility and Statutory Rigidity*, 19 U. MICH. J.L. REF. 109 (1985).

[82] *See generally* Schaffer & Fox, *supra* note 81; Irish & Cohen, *supra* note 81; Shaw v. Delta Air Lines, Inc., 463 U.S. 85, 98 (1983).

sons acting on behalf of" ERISA-covered plans, and state laws that "relate to" funding and benefits-vesting provisions of pension plans.[83] The Senate version would have preempted state laws that "relate to the subject matters regulated by this Act."[84]

In the final joint conference on the bills, the Conference Committee abandoned these approaches and adopted the present language. When the Conference Committee Report was made available to the full Congress only ten days before the bill was enacted, little was said about the change.[85] The Committee Report said nothing about the change, other than describing the provision.[86] Senator John Williams (R-Del.) told the Senate that the broad language would make it impossible for "state professional associations" to prevent "unions and employers" from agreeing on particular benefit plans.[87] The concern was whether ERISA would prevent state bar associations from prohibiting "closed panel" prepaid legal services plans.[88] Section 514 was broadened to preclude such enforcement of state legal ethics rules against plans.[89] The problem with the narrower language,

---

[83] H.R. 2, 93d Cong., 1st Sess. (1973), 120 CONG. REC. 4742 (1974).

[84] S. 4200, 93d Cong., 1st Sess. (1973), 120 CONG. REC. 5002 (1974).

[85] *See* Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724, 745 n.23 (1985).

[86] H.R. CONF. REP. No. 1280, 93d Cong., 1st Sess. (1973); S. CONF. REP. No. 1090, 93d Cong., 1st Sess. (1973).

[87] 120 CONG. REC. 29,933 (1974) (remarks of Sen. Williams (R-Del.)). Courts often cite Sen. Williams' floor statement and a similar one by Representative Dent in the House as evidence of congressional intent to preempt broadly. 120 CONG. REC. 29,197 (1974) (remarks of Rep. John Dent (D-Pa.)). *See, e.g.*, Shaw v. Delta Air Lines, Inc., 463 U.S. 85, 99 (1983). As principal sponsors of the bills in the Senate and House, their statements were given great weight when the Court first defined the contours of ERISA preemption, during the high point of the use of legislative history. Without entering the debate over whether such statements are reliable, I would simply note that the statements, like the language of the statute, are ambiguous.

[88] Specifically, the question was whether ERISA would preempt a state legal ethics law regulating whether an employer-provided legal services plan could limit the participants' choice of lawyers, much as an HMO limits the choice of doctors. *See* Robert S. McDonough, Note, *ERISA Preemption of State Mandated-Provider Laws*, 1985 DUKE L.J. 1194, 1201.

[89] The interest group politics surrounding preemption involved more than simply the question of whether state bars could regulate prepaid legal services plans. Organized labor apparently wanted ERISA to preempt state health laws mandating benefits in order to prevent such laws from circumscribing their freedom in collective bargaining. Schaffer & Fox, *supra* note 81, at 51. Labor's chief lobbyist on ERISA is reported to have said in 1987, "We understood we were giving up good state mandated benefits but we wanted the freedom to give up particular benefits in return for cash wages, and to trade in one benefit for another." *Id.* Apparently business remained relatively quiet on this aspect of preemption, and the insurance industry remained silent on almost all aspects. *Id.*

Other provisions of ERISA's preemption provision also were added to fix specific perceived problems without apparent awareness of the possible ramifications of the language. The "deemer clause," which prevents states from regulating self-insured

Senator Williams suggested, was that it would have given rise to difficult line-drawing problems.[90] In a futile effort to save ERISA from the line-drawing that dogs every preemption issue, a frustrated conference committee decided at the last minute to preempt "any and all" laws that "relate to" plans covered by ERISA. Congress wanted to enact legislation, and it fixed the immediate problem with the bills it had before it.

This legislative history confirms the observation that plain language interpretation and the actual legislative process of writing statutory language work at cross purposes.[91] Drafters tend to choose language to fit a paradigmatic case and then try to imagine circumstances where the language might produce an undesirable result. ERISA shows how their imaginations can be constrained by a shortage of time or experience. Drafters also tend to think about the meaning of a particular provision in the context of the entire statutory scheme, which courts often fail to do.[92] And, as the difficulty of revising ERISA's preemption language suggests, legislatures always face collective action problems that make the renegotiation of language costly, difficult, and unpredictable. In the context of an enormous and complex statute, those problems may become especially acute.[93] The inferences that courts have drawn from the expansion of ERISA's preemption language are the wrong inferences.

The Act and its legislative history both suggest that those members of Congress who paid attention to the preemption issue may have thought that the quick fix to the preemption problem was provisional. Section 3022 mandated the creation of a Joint Pension Task Force to study the practical effect and desirability of federal preemption. In addition, in commenting on the newly broadened preemption provision, Senator Javits (D-Fla.) said that "the desirability of further regulation—at either the State or

---

ERISA plans, *see infra* text accompanying notes 157–167, was a response to a lower court decision in Missouri that treated a benefit plan as an insurance company and reportedly fined Monsanto $185 million for operating an insurance company without a license. *See* Conison, *supra* note 27, at 648–49.

[90] 120 CONG. REC. 29,197 (1974) (remarks of Sen. Williams).

[91] *See* Eric Schnapper, *Statutory Misinterpretations: A Legal Autopsy*, 68 NOTRE DAME L. REV. 1095, 1107–08 (1993) (making this observation with regard to the Civil Rights Act of 1991); *cf.* James J. Brudney, *Congressional Commentary on Judicial Interpretations of Statutes: Idle Chatter or Telling Response?*, 93 MICH. L. REV. 1 (1994).

[92] Farber, *supra* note 79, at 550–52.

[93] Mathew McCubbins et al., *Legislative Intent: The Use of Positive Political Theory in Statutory Interpretation*, 57 LAW & CONTEMP. PROBS. 3, 14 (1994).

Federal level—undoubtedly warrants further attention."[94] But the Task Force study never materialized. Although Congress clearly thought that its preemption solution was provisional, it is not clear what role Congress intended courts to play in fine-tuning preemption, or how Congress's apparent desire to return to the problem affects the latitude it expected courts to exercise. What is clear, however, is that to the extent Congress thought at all about preemption, "section 514(a) was more in the nature of a quick statement of general principle than a workable, final rule."[95] Thus, "neither before nor after enactment of ERISA did Congress . . . view preemption policy issues as settled in the way that the sweeping statutory language of section 514 might suggest."[96] The change to section 514 is simply not evidence that Congress intended the enormously broad preemption, especially of laws remote from ERISA's purposes, that courts have created.[97]

Thus, the problem in interpreting section 514 is not simply that the plain language is unhelpful; other forms of what I have termed textualism are equally unavailing. If one looks to the purpose of the statute, one could concoct an argument either way as to whether its protective purposes would be better served by compelling national uniformity on all these matters (even if the nationwide standard is one of no regulation) or by allowing states to enforce legislation for the benefit of employees who are also the supposed beneficiaries of ERISA's protections. Nor is a consideration of the structure of the Act much more enlightening. That ERISA itself does not regulate the terms of employment in apprenticeship programs could be viewed as evidence that it allows the states to play their traditional role in regulating such programs. On the other hand, that ERISA does not regulate the terms of health benefit plans is generally not taken to mean that states could play their traditional role in regulating those conditions of employment. Again, the complexity of the legislative process makes it difficult to infer from ERISA's structure and coverage exactly what role Congress intended state law to

---

[94] 120 CONG. REC. 29,942 (1974) (remarks of Sen. Jacob Javits).

[95] Irish & Cohen, *supra* note 81, at 114.

[96] *Id.* at 116.

[97] Irish & Cohen observe that Congress enacted ERISA, including its preemption provision, while it was ignorant of the full range and complexity of the issues surrounding employee benefits. *Id.* at 11. The authors also note that "[a]t least one of ERISA's principal authors has consistently suggested that the apparent principle section 514(a) states is broader than the rule that ought to be enforced." *Id.*

play. But certainly the Court has erred in seeing ERISA's apparent comprehensiveness as evidence of an intent to supersede all state law.[98]

In sum, the legislative history of ERISA suggests at least three problems with a preemption inquiry that is a search for congressional intent.[99] One is that, as I have just described, intent simply did not exist on much of anything beyond the desire to federalize an area of law that previously had been left to the states while avoiding line-drawing problems of the sort exemplified by the prepaid legal services plan dispute. A second problem stems from the fact that Congress did not give a great deal of thought to whether the scope of preemption should reflect the different degrees of federal regulation of pension plans, as opposed to welfare benefit plans. Broad preemption of state law may make sense when Congress decides to regulate a field extensively, as it did with respect to pensions. But broad preemption makes little sense when Congress does not extensively regulate in an area, as is the case with nonpension benefits. There is no evidence that Congress realized that broad preemption of state law would create a large regulatory void with regard to nonpension benefits in particular, nor is there evidence that employers, plans, and insurance companies realized they could use ERISA preemption as a shield against a very wide range of state regulation.

The third difficulty with the conceptualization of preemption as a search for congressional intent is that times have changed so much that the meaning of Congress's choice to preempt broadly is drastically different today, when the social context relevant to state and federal regulation of employee benefits differs dramatically from what it was in 1974. The changed context means that giving effect to Congress's particularized intent—broad preemption—thwarts Congress's general intent—creation of a compre-

---

[98] *See, e.g.*, Shaw v. Delta Air Lines, Inc., 463 U.S. 85, 98 (1983); Pilot Life Ins. Co. v. Dedeaux, 481 U.S. 41, 54 (1987). In another context, commentators have cautioned against attributing too much comprehensiveness to any statute:

> Public choice teaches that a statute reflects not only the preferences of the legislature, but also the procedural obstacle course of enactment. The fact that a statute explicitly regulates situations A and B, but not C, should not necessarily be interpreted as a decision to immunize C from regulation. It may only indicate that, for whatever reason, the legislative process failed to produce a bill covering C.

Daniel A. Farber & Philip P. Frickey, *In the Shadow of the Legislature: The Common Law in the Age of the New Public Law*, 89 MICH. L. REV. 875, 892 (1991) (footnote omitted).

[99] Although I limit my conclusions to ERISA preemption, the same concerns may also apply beyond the scope of ERISA.

hensive federal system to protect beneficiaries of plans. One of the major changes is the shift of responsibility for health and welfare issues from the federal government to state and local governments, which began during the administrations of Presidents Ford and Carter but reached a frantic pace during the Reagan Administration and continues now in the 104th Congress. From the late 1970s to the present, with the exception of the first two years of the Clinton Administration, political efforts to expand access to health care increasingly have focused on the states. In 1974, "Senators Javits and Williams would have had reasonable grounds to believe that federal law, perhaps even national health insurance, would fill the regulatory gap that ERISA preemption had created."[100] ERISA's broad preemption of state law may have been regarded simply as a prelude to the creation of a national social insurance scheme for both health and pensions. Congress wanted to make private social insurance a purely federal concern in order to control the integration of private plans with Social Security and a national health insurance program.

However, after 1974, the movement for national social insurance lost steam. ERISA's nationalization of the law regulating privately provided social insurance was not followed by a national system addressing the social insurance needs that ERISA's private-contract approach left unmet. Because states and cities could not afford to provide the social insurance that the federal government would not, the effort to shift those costs to business became politically viable. The rising cost of health care only intensified the pressure. Yet preemption has prevented state efforts to solve the health insurance crisis by imposing costs on business. The problem of ERISA preemption has become acute because of institutional changes in federalism and social insurance that Congress did not anticipate in 1974.

In sum, section 514 is ambiguous. Courts must make significant choices about the scope of ERISA's preemptive effect; the language, legislative history, and purpose of the statute simply do not dictate answers.

---

[100] Schaffer & Fox, *supra* note 81, at 53.

II. The Tortured History and an Explanation of ERISA Preemption

Now that the Supreme Court appears to be entering a new era in its interpretation of the scope of ERISA preemption, it is time to examine what led to the failure of the old approach. In this Part, I will show how the Court relied on three variations of textualism in interpreting the ERISA preemption provision and why its reliance was misplaced. The Court's plain language approach could not decide difficult cases, which the Court ultimately admitted.[101] The Court's reliance on congressional intent hardly fared better, because the legislative history shows that the preemption language was an eleventh-hour fix to a particular problem rather than a considered choice about federalism in the full range of subjects that are "related to" employee benefits.[102] Finally, the inconsistency in the Court's treatment of statutory purpose reveals that reliance on legislative "purpose" provided little more guidance than reliance on language or legislative history.[103] In short, all three versions of the Court's textualism flopped.

The Court's methods neither provided certainty to the law nor absolved the Court of the responsibility for defining the relationship between state and federal law. The Court's failure to develop a coherent approach to preemption generated uncertainty, and its overbroad plain language analyses led to challenges to almost every kind of state regulation having an impact on employee benefit plans. In the twenty-one years since ERISA was enacted, the Court has rendered decisions with written opinions in twelve ERISA preemption cases,[104] and has decided a number

---

[101] *See infra* text accompanying notes 123–142; District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580 (1992); *New York Blues*, 115 S. Ct. 1671 (1995).

[102] *See infra* text accompanying notes 143–154.

[103] *See infra* text accompanying notes 171–188.

[104] *New York Blues*, 115 S. Ct. 1671 (1995); John Hancock Mut. Life Ins. Co. v. Harris Trust & Sav. Bank, 114 S. Ct. 517 (1993); *Greater Washington*, 113 S. Ct. 580 (1992); Ingersoll-Rand Co. v. McClendon, 498 U.S. 133 (1990); FMC Corp. v. Holliday, 498 U.S. 52 (1990); Massachusetts v. Morash, 490 U.S. 107 (1989); Mackey v. Lanier Collection Agency & Serv., Inc., 486 U.S. 825 (1988); Fort Halifax Packing Co. v. Coyne, 482 U.S. 1 (1987); Pilot Life Ins. Co. v. Dedeaux, 481 U.S. 41 (1987); Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724 (1985); Shaw v. Delta Air Lines, Inc., 463 U.S. 85 (1983); Alessi v. Raybestos-Manhattan, Inc., 451 U.S. 504 (1981).

Two other cases address ERISA preemption in determining the limits of removal jurisdiction over cases originally filed in state court and then removed to federal court

of others without opinion.[105] Preemption cases constitute roughly half of all the ERISA cases the Court has considered. The relatively large number of ERISA preemption opinions has not, however, led to clarity in the law. The lower courts have decided thousands of preemption cases,[106] yet remain mired in confusion about basic points. ERISA preemption offers proof that plain language textualism leads to uncertainty and incoherence in the law.[107] Moreover, although the Court relied on textualism to avoid the responsibility for deciding the appropriate balance of state and federal regulation, asserting that the irrational law it created was the fault of Congress,[108] the Court simply deluded itself about its responsibility for devising a coherent body of law.

The Court's primary preference was for a rule of interpretation (textualism) that seemed ideologically neutral; this preference explains the large number of unanimous opinions on issues that one would not expect to produce unanimity in an ideologically divided Court. Textualism was particularly appealing to a Court confronting a complex statute in an unfamiliar field of law. In this sense, faith in textualism was both cause and effect—it was a partial cause of the Court's failure to appreciate the implications of its decisions, but it was also an effect of the Court's lack of vision.

Textualism was not the only invisible agent. The ERISA cases also reveal the Court's historic preference for national rather than state control of labor law. The Court applied the framework of national dominance that it had developed for labor law to the new law governing employee benefit plans, even where federal law amounted to a preference for total employer discretion in designing social insurance arrangements free of governmental

---

on the basis of complete preemption. Metropolitan Life Ins. Co. v. Taylor, 481 U.S. 58 (1987); Franchise Tax Bd. v. Construction Laborers Vacation Trust, 463 U.S. 1 (1983).

[105] Local Union 598, Plumbers & Pipefitters Indus. Journeymen & Apprentices Training Fund v. J.A. Jones Constr. Co., 846 F.2d 1213 (9th Cir. 1991), *aff'd*, 488 U.S. 881 (1988); Stone & Webster Eng'g Corp. v. Ilsley, 690 F.2d 323 (2d Cir. 1982), *aff'd sub nom.* Arcudi v. Stone & Webster Eng'g Corp., 463 U.S. 1220 (1983); Standard Oil Co. v. Agsalud, 633 F.2d 760 (9th Cir. 1980), *aff'd*, 454 U.S. 801 (1981).

[106] A recent search in the Westlaw "Allfeds" database (ERISA /p preempt!) produced 3330 cases. In a 1992 Lexis search, Justice Stevens found over 2800 judicial opinions addressing ERISA preemption. *Greater Washington*, 113 S. Ct. at 586 n.3 (Stevens, J., dissenting).

[107] *See* Pierce, *supra* note 21.

[108] *See* FMC Corp. v. Holliday, 498 U.S. 52, 62 (1990) ("[W]e merely give life to a distinction created by Congress . . . ." (quoting Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724, 747 (1985))).

mandate. Under ERISA, as under the National Labor Relations Act, the consensus favoring broad preemption did not begin to erode until after the serious political consequences of broad preemption of state law became apparent. But because the Court had said that broad ERISA preemption was dictated by the language of the statute, revision of ERISA preemption doctrine was more difficult than revision of the more flexible implied preemption doctrine of the NLRA.

A. *The Evolution of ERISA Preemption in the Supreme Court*

In this section, I will explore the effect of the Court's interpretive practices on the development of ERISA preemption doctrine.[109] The development of the law has not been orderly. At times the Court has focused quite rigidly on plain language, while at other times it has strained against its plain language precedents to reach results that seem more sensible. Whether relying on the text of ERISA's preemption provision or relying on its purpose, the Court has obscured the value choices it has made about employee benefits, but it certainly has not avoided making choices.

1. The Rise and Fall of Plain Language

The Court's initial approach to ERISA preemption combined textualism with a substantive vision of national dominance borrowed from the National Labor Relations Act. In its first case considering preemption, *Alessi v. Raybestos-Manhattan, Inc.*,[110] a unanimous Court, per Justice Marshall, held that ERISA preempted a New Jersey workers' compensation law prohibiting pension plan provisions which deducted workers' compensation benefits from pension benefits.[111] The New Jersey law was designed to prevent employers from structuring their plans so that workers'

---

[109] For another example of a case study on the effect of an interpretive method across an entire substantive area of law, see Robert K. Rasmussen, *A Study of the Costs and Benefits of Textualism: The Supreme Court's Bankruptcy Cases*, 71 WASH. U. L.Q. 535 (1993).

[110] 451 U.S. 504 (1981).

[111] *Id.* at 506. In invalidating the state law, the Court accepted the position of the employer, who was joined by the United States, the Chamber of Commerce, various corporations, and the ERISA Industry Committee as *amici*.

pension benefits would be reduced when they received workers' compensation.

The Court began its preemption analysis by invoking "respect for the separate spheres of governmental authority preserved in our federalist system."[112] The New Jersey statute governed state workers' compensation awards, which, the Court acknowledged, are an area of traditional state concern "obviously . . . subject to the State's police power."[113]

Notwithstanding the invocation of federalism, the Court cut a wide swath for federal law. The Court initially observed that the inclusion of an explicit preemption provision in ERISA made it plain that Congress "meant to establish pension plan regulation as exclusively a federal concern."[114] Yet, the Court noted, the "relates to" language of section 514 "gives rise to some confusion where, as here, it is asserted to apply to a state law ostensibly regulating a matter quite different from pension plans."[115] The state law was presumably intended to protect workers' rights to their workers' compensation awards,[116] and thus was not inconsistent with the purpose or requirements of ERISA. The Court nevertheless concluded that "[w]hatever the purpose or purposes of the New Jersey statute, we conclude that it 'relate[s] to pension plans' governed by ERISA because it eliminates one method for calculating pension benefits—integration—that is permitted by federal law."[117] Thus, without much discussion, the Court decided that ERISA preemption is broader than ordinary federal preemption, which displaces only laws inconsistent with provisions or goals of federal law or in areas that federal law regulates.[118] The Court also determined that the subject and purpose of a state law are irrelevant to the ERISA preemption inquiry. These decisions turned out to be very important for the later development of ERISA preemption doctrine.

The Court's policy justification for its ruling was a vision that ERISA protects the "rights" of plan designers to structure their

---

[112] *Id.* at 522.
[113] *Id.* at 524.
[114] *Id.* at 523.
[115] *Id.* at 523–24.
[116] *Id.* at 524.
[117] *Id.* at 524. This rationale—analogizing integration of workers' compensation benefits to integration of Social Security benefits—is troubling on the merits. The fact that ERISA permits integration with Social Security says nothing about whether pensions ought to be reduced due to receipt of an entirely different kind of benefit.
[118] *See supra* text accompanying notes 30–39.

plans as they see fit: "ERISA leaves integration, along with other pension calculation techniques, subject to the discretion of pension plan designers. Where, as here, the pension plans emerge from collective bargaining, the additional federal interest in precluding state interference with labor-management negotiations calls for preemption of state efforts to regulate pension terms."[119] This rationale placed the Court's ERISA preemption cases within the same conceptual framework it had used in determining the scope of preemption implied by the National Labor Relations Act—state laws that constrain the discretion of labor and management are preempted because they interfere with the regime of collective bargaining. This is consistent with what the Court had done in its line of labor preemption cases that began with *Lodge 76, International Association of Machinists v. Wisconsin Employment Relations Commission*.[120] The Court apparently overlooked the problem it created by applying to ERISA (under which most plans are *not* collectively bargained) a laissez-faire policy which makes sense only because it assumes that employees are protected by collective bargaining.

The invocation of an asserted "federal interest in precluding state interference with labor-management negotiations"[121] was a mistake insofar as it suggested that state laws guaranteeing minimum working conditions would be preempted because they interfere with labor-management negotiations. In later cases, where employers invoked this rationale to seek ERISA preemption of a variety of state protective labor laws, the Court backed away from it.[122] *Alessi*'s notion that plan design should be left to "the discretion of pension plan designers" conflated a vaguely articulated ERISA policy favoring national uniformity with a policy favoring unrestricted discretion in setting terms of employment, even though the latter was *not* part of ERISA and was only somewhat a part of the NLRA. In this way, the Court unwittingly

---

[119] Alessi v. Raybestos-Manhattan, Inc., 451 U.S. 504, 525 (1981).

[120] 427 U.S. 132 (1976). The essence of the *Machinists* preemption doctrine is the notion that Congress intended some conduct that the NLRA neither protects nor prohibits to be left entirely unregulated.

[121] *Alessi*, 451 U.S. at 525.

[122] In Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724 (1985); Fort Halifax Packing Co. v. Coyne, 482 U.S. 1 (1987); and Massachusetts v. Morash, 490 U.S. 107 (1989), the Court upheld state insurance regulation and minimum working condition legislation against preemption challenges even though the legislation effectively regulated terms of employee benefit plans.

transformed a preference for national uniformity into a preference for the absence of regulation altogether.

The Court's preference for a textualist approach to ERISA preemption became even more apparent in its second published ERISA preemption decision,[123] which, like *Alessi*, also invalidated a state protective labor law. In *Shaw v. Delta Air Lines, Inc.*,[124] a unanimous Court held that ERISA preempts two state disability and human rights laws that prohibited discrimination on the grounds of pregnancy and required employers to provide sick leave to employees disabled by pregnancy.[125] Justice Blackmun wrote for the Court that the New York Human Rights Law, "which prohibits employers from structuring their employee benefit plans in a manner that discriminates on the basis of pregnancy," and the New York Disability Benefits Law, "which requires employers to pay employees specific benefits, clearly 'relate to' benefit plans."[126]

---

[123]The Court's second ERISA preemption case did not produce a written opinion. In Standard Oil Co. v. Agsalud, 633 F.2d 760 (9th Cir. 1980), *aff'd*, 454 U.S. 801 (1981), the Court summarily affirmed the Ninth Circuit's holding that ERISA preempts a Hawaii law requiring employers to provide certain health care benefits for their employees. This holding was partially overturned when Congress amended the preemption provision to save part of Hawaii's statute. 29 U.S.C. § 1144(b)(5)(A) (1988).

The Court also used the summary procedure to establish the contours of preemption in another early case, Stone & Webster Engineering Corp. v. Ilsley, 690 F.2d 323 (2d Cir. 1982), *aff'd sub nom.* Arcudi v. Stone & Webster Engineering Corp., 463 U.S. 1220 (1983). The Court invalidated a Connecticut statute that required employers to continue health, accident and life insurance coverage for their employees while they received workers' compensation benefits.

Connecticut later successfully skirted the *Stone & Webster* obstacle by amending its workers' compensation statute to require employers to provide coverage, thus squeezing protection into the § 4(b)(3) exception for plans maintained to comply with workers compensation statutes. The Second Circuit rejected a preemption challenge to the revised statute. R.R Donnelley & Sons Co. v. Prevost, 915 F.2d 787 (2d Cir. 1990), *cert. denied*, 499 U.S. 947 (1991). The Supreme Court disapproved the *Donnelley* result and invalidated a scheme like Connecticut's in District of Columbia v. Greater Washington Bd. of Trade, 113 S. Ct. 580, 583 (1992).

[124]463 U.S. 85 (1983).

[125]The Court held that ERISA preempted the laws only to the extent that their protections were more generous than Title VII (which the Court had previously held not to prohibit discrimination on the basis of pregnancy, General Electric Co. v. Gilbert, 429 U.S. 125, 133–46 (1976)). In other words, the Court held that employee benefit plans must comply with Title VII's antidiscrimination provisions but need not comply with state antidiscrimination laws. Since Title VII did not then prohibit pregnancy discrimination, the effect of the decision was to insulate plans from more egalitarian state antidiscrimination laws. The reason for the partial nonpreemption is that ERISA does not invalidate other federal laws, and Title VII relies on state law to enforce some of its protections.

[126]*Shaw*, 463 U.S. at 97 (1983). The actual holding of *Shaw* was a bit more complicated; its very complexity illustrates the irrationality of ERISA preemption. The Court held that the state disability law, which required employers to pay benefits to disabled employees equal to half the average weekly wage, was preempted as applied

In *Shaw*, the Court passed up the chance to distinguish preemption in the pension area from preemption in the non-pension area by failing to consider the possibility that the preemption clause could be interpreted differently in the two circumstances. The Court asserted as a matter of course that the meaning of the term "relates to" is unambiguous and that it necessarily requires broad preemption: "A law 'relates to' an employee benefit plan, in the normal sense of the phrase, *if it has a connection with or reference to* such a plan."[127] As support for this proposition, the Court quoted the *Black's Law Dictionary* definition of "relate."[128] The Court also concluded that it was necessary to read "relates to" in its broadest sense, because it would otherwise have been unnecessary to exempt generally applicable state criminal laws from preemption, as section 514 does.[129]

The Court's reliance on the dictionary language "connection with or reference to" did nothing to reduce uncertainty about the scope of preemption, but instead created the possibility of exceedingly broad preemption. Just as the term "relates to" requires some modifier in order to have any useful meaning, so too do the terms "connection with" and "refer to," which the Court used to explain the meaning of "relate to." Although the Court implicitly recognized the potentially unlimited reach of "relates to" by creating in a footnote an exception that some state laws "may affect employee benefit plans in too tenuous, remote, or peripheral a manner to warrant a finding that the law 'relates to' the plan,"[130] the Court did not define the scope of the limit, nor did it try to derive it from the language, legislative history, or purpose of ERISA.[131] The Court's insistence that broad

---

to plans that were not maintained as a separate administrative unit solely to comply with the state law. If the employer complied with the law by including state-mandated disability benefits along with other benefits, the plan would be covered by ERISA and the state law would be unenforceable against it. *Id.* at 107–08. It is difficult to see what purpose is served by allowing enforcement of state law against some disability plans but not others. The decision rests on the exemption from ERISA coverage of plans maintained *solely* to comply with state disability insurance laws. 29 U.S.C. § 1003(b)(3) (1988). The Court's wooden reading of the word "solely" converted a provision likely intended to prevent evasion of ERISA's requirements into a provision that allows evasion of state law where the employer is not required by state law to maintain a separate plan.

[127] *Shaw*, 463 U.S. at 96–97 (emphasis added).

[128] "To stand in some relation; to have bearing or concern; to pertain; refer; to bring into association with or connection with." *Id.* at 97 n.16.

[129] *Id.* at 98.

[130] *Id.* at 100 n.21.

[131] However, it did cite a Second Circuit decision finding an "implied exception" to ERISA preemption for state domestic relations orders. *Id.* at 100 n.21 (quoting

preemption was mandated by the "relates to" language, combined with its contrary invention of the "tenuous, remote, or peripheral" exception, revealed the contradictions in both the holding and the reasoning of the case.

In *Shaw*, the Court also resorted to the legislative history of ERISA to support its reading of the Act's language.[132] The Court surmised that the existence of specific exceptions to preemption and the rejection of narrower versions of the preemption provision during the legislative process made it clear that Congress "used the words 'relate to' in § 514(a) in their broad sense."[133] The Court quoted Senator Javits's remarks on the floor of the Senate, in which the Senator explained that broad preemption language was substituted for narrower language in the Senate and House bills because the bills that related preemption to the areas of federal regulation "raised the possibility of endless litigation over the validity of State action that might impinge on Federal regulation."[134] In choosing broad preemption, the Court gave effect to a congressional decision to preempt more state laws than those dealing with subjects to which ERISA directly spoke (i.e., pension law and trust law). But there is no evidence that Congress had considered how far beyond the substantive perimeters of ERISA it intended to preempt state law.

Although *Shaw* was the beginning of the Court's fruitless pursuit of a plain language approach to ERISA preemption, the Court did not rely entirely on ERISA's language until the plain language approach reached its zenith (or nadir) in *District of Columbia v. Greater Washington Board of Trade*,[135] a case in which the Court found "linguistic precision where it does not exist."[136] Justice Thomas, writing for eight members of the Court, took the Court's broadest position yet in asserting that ERISA preempts any state law that "*refers to* welfare benefit plans" or that "im-

---

American Tel. & Tel. Co. v. Merry, 592 F.2d 118, 121 (2d Cir. 1979)). This may have been intended to signal a limit on preemption. The Court had previously signaled such a limitation by dismissing for want of a substantial federal question appeals of lower court decisions holding that ERISA does not preempt court-ordered spousal support to divorced nonemployee spouses. Carpenters Pension Trust Fund v. Campa, 444 U.S. 1028 (1980).

[132] In recent times, however, the Court has rarely refered to legislative history. *See, e.g.*, Thomas W. Merrill, *Textualism and the Future of the Chevron Doctrine*, 72 WASH. U. L.Q. 351, 355 (1994).

[133] *Shaw*, 463 U.S. at 98 (1983).

[134] *Id.* at 99 n.20 (quoting 120 CONG. REC. 29,942 (1974) (remarks of Sen. Javits)).

[135] 113 S. Ct. 580 (1992).

[136] Pierce, *supra* note 21, at 752. This case is a stark example of what Pierce calls the Court's "hypertextualism." *Id.*

pos[es] . . . requirements *by reference to* such covered programs."[137] At issue was a provision of the District of Columbia's workers' compensation statute which required any employer who provided health insurance for its employees to continue such health coverage while employees received workers' compensation benefits. Relying on the *Black's Law Dictionary* definition of "relates to" and what Justice Thomas characterized as the "ordinary meaning" of the term, the Court found "relates to" to be synonymous with "refers to" or "makes reference to."[138] The Court held the workers' compensation provision preempted because it set compensation benefit levels by reference to the amount of employee benefits employers pay.[139]

The outcome of *Greater Washington* is at least arguably defensible. One could conceivably read ERISA as preempting state laws imposing substantial social insurance costs on employers with benefit plans while not imposing costs on employers without plans. What is preposterous about the case is its reasoning. There is no reason to believe that any state law that simply *mentions* an ERISA plan should be preempted. As Justice Stevens pointed out in dissent, the Court's opinion calls into question ordinary principles of the state laws of tort and contract damages, which compute a wrongfully terminated or injured employee's lost wages by adding to the take-home pay the value of fringe benefits such as health insurance and vacation pay.[140] Justice Stevens's example turned out not to be fanciful, as employers relied on the majority opinion to challenge state laws that

---

[137] 113 S. Ct. at 583, 584 (emphases added).

[138] *Id.* at 583.

[139] *Id.* at 584. The Court rejected D.C.'s argument that the statute should be saved from preemption even if it "related to" an ERISA plan because the same result could be achieved by requiring employers to maintain separate workers' compensation benefit plans exempt from ERISA regulation under § (b)(3), 29 U.S.C. § 1003(b). Characterizing this interpretation as a "two-step analysis," Justice Thomas rejected the contention summarily: "We cannot engraft a two-step analysis onto a one-step statute." *Greater Washington*, 113 S. Ct. at 585. He did not explain what makes § 514(a) a "one-step statute."

[140] *Id.* at 585. As Justice Stevens observed in a footnote, this reading of ERISA preemption had been considered and rejected in several lower court cases. *See id.* at n.1 (citing Martori Bros. Distributors v. James-Massengale, 781 F.2d 1349, 1358–59 (9th Cir. 1986), *modified*, 791 F.2d 799 (9th Cir. 1986), *cert. denied*, 479 U.S. 949 (1986); Teper v. Park West Galleries, Inc., 427 N.W.2d 535, 541 (Mich. 1988); Schultz v. National Coalition of Hispanic Mental Health and Human Services Organizations, 678 F. Supp. 936, 938 (D.D.C. 1988); Jaskilka v. Carpenter Technology Corp., 757 F. Supp. 175, 178 (D. Conn. 1991). *See also* Ethridge v. Harbor House Restaurant, 861 F.2d 1389, 1405 (9th Cir. 1988) (holding that ERISA does not preempt wrongful discharge claims in which damages include loss of future benefits, if discharge was not motivated by desire to avoid paying benefits).

simply mentioned ERISA plans. For instance, the Court's language generated a challenge to Michigan's business tax on the value added to goods, including the cost of labor (which is measured by wages and benefit plan contributions made).[141]

It is difficult to believe that the Court actually meant what the opinion says, i.e., that a state law that "specifically refers to welfare benefit plans regulated by ERISA [is] on that basis alone" preempted.[142] The opinion illustrates in extreme form the failure of "plain language" textualism as a device for dealing with ERISA preemption: The language of the preemption provision is fundamentally ambiguous. A state law may "relate to" an employee benefit plan to a very slight extent or to a very great extent, and the "tenuous, remote, peripheral" exception neither provides clarity nor has been given any teeth by the Court. The result has been indeterminacy in the law and understandable confusion in the lower courts.

2. In the Pursuit of Legislative Intent and Statutory Purpose

In other cases, the Court blended a plain language approach with reliance on legislative intent and statutory purpose, but the language as the Court defined it was in tension with Congress's apparent intent and purpose to protect employees. To reconcile the apparent conflict, the Court redefined ERISA as protecting employees by promoting national uniformity and administrative efficiency through the elimination of state regulation. For instance, the Court's decision in *Metropolitan Life Insurance Co. v. Massachusetts*[143] was the first of two cases that rejected ERISA preemption of state laws mandating minimum terms in insurance policies, but held that such laws could *not* be applied to self-insured plans.[144] The Court mixed a plain language analysis with a purposive analysis in an effort to reconcile ERISA's protective purposes with what the Court thought to be sweeping preemption of state law. The Court analyzed the language of the insurance savings clause (which saves from preemption state laws that "regulate insurance") and the exception to it (the so-called

---

[141] The court rejected the challenge. *See* Thiokol Corp. v. Roberts, 858 F. Supp. 674 (W.D. Mich. 1994).

[142] *Greater Washington*, 113 S. Ct. at 583.

[143] 471 U.S. 724 (1985).

[144] The second such case was FMC Corp. v. Holliday, 498 U.S. 356 (1990), which is discussed *infra* at notes 155–167.

"deemer clause," which states that plans may not be deemed to be insurance for the purpose of the insurance savings provision) to create a result that even the Court had to admit did not make much sense.

A unanimous Court, per Justice Blackmun, held that a Massachusetts law requiring that certain health care benefits be included in any health insurance policy or employee health benefit plan was not preempted as applied to insurance policies because it was saved by the state insurance law exception. However, the provision could not be applied to benefit plans directly, because to the extent it was applied directly it was not an insurance law and hence was not saved. The Court thus created a framework that enabled employee benefit plan sponsors to evade state regulation by self-insuring rather than purchasing insurance. The Court recognized that allowing states to regulate insured plans but not self-insured plans thwarted the alleged purpose of creating uniform national law, at least for plans that purchase insurance, and the Court also conceded that the differential regulation of insured and uninsured plans served no useful purpose and was probably unintended.[145] Yet, the Court disclaimed responsibility for the irrational result; it stated, "we merely give life to a distinction created by Congress," and pointed out that a congressional committee had become aware of the problem some years after ERISA was enacted but that legislation to correct the problem had died in the Senate.[146] The Court's distinction between insured and self-insured plans led employers to self-insure to avoid state regulation, which in turn has led to a significant but unintended shift in the structuring and financing of health plans.[147]

Perhaps recognizing the weaknesses of a plain language approach that produces irrational law, the Court looked to the purpose of the statute for additional support. In rejecting the contention that ERISA ought not preempt state mandated-benefits laws that both concern subjects that ERISA does not regulate and are consistent with ERISA's protective purposes, the Court

---

[145] *Metropolitan Life Insurance Co.*, 471 U.S. at 747.

[146] *Id.* at 747 n.25.

[147] A recent report by the General Accounting Office concludes that self-funding of health plans has increased among both large and small businesses, and notes that state officials fear that the increase of self-funding poses a danger to plan beneficiaries because self-funded plans may be inadequately funded. HEALTH, EDUCATION, AND HUMAN SERVICES DIVISION, U.S. GAO, EMPLOYER-BASED HEALTH PLANS: ISSUES, TRENDS, AND CHALLENGES POSED BY ERISA, REP. NO. GAO/HEHS 95-167 (July 1995). *See infra* text accompanying notes 168–170.

rejected two possible limits on preemption. After noting again the "broad scope" of the preemption clause, the Court asserted that the clause was "intended to displace all state laws that fall within its sphere, even including state laws that are consistent with ERISA's substantive requirements."[148] In the course of its discussion, the Court returned to the difficult problem of discerning Congress's intent. It noted that the significant broadening of the preemption provision happened "at the last minute," that it was not carefully considered, and that it was broadened for a rather narrow reason.[149] Yet, notwithstanding whatever doubt the Court harbored about the rationality or perhaps even the clarity of Congress's intent in adopting the "relate to" language, the Court did not back away from its prior broad preemption holdings.

The Court did, however, back away from the rationale for broad preemption that it had hinted at in *Alessi v. Raybestos-Manhattan, Inc.*[150] In particular, the Court qualified the notion of national laissez-faire in employee benefits which *Alessi* had seemed to invoke. It did so not in construing ERISA preemption, but rather in rejecting the insurer's argument that the mandated benefit law was preempted by the National Labor Relations Act. The argument was that state laws setting minimum terms of employment interfere with collective bargaining, and therefore are preempted by the NLRA.[151] The Court reasoned that "[t]he NLRA is concerned primarily with establishing an equitable process for determining terms and conditions of employment, and not with particular substantive terms of the bargain that is struck."[152] Further, the Court noted, "[m]inimum state labor standards affect union and nonunion employees equally, and neither encourage nor discourage the collective-bargaining processes that are the subject of the NLRA."[153] Most important, the Court sug-

---

[148] *Metropolitan Life Insurance Co.*, 471 U.S. at 739.

[149] *Id.* at 745 n.23.

[150] 451 U.S. 504 (1981).

[151] This is the branch of labor law preemption doctrine known as "*Machinists* preemption" (after Lodge 76, International Association of Machinists & Aerospace Workers, AFL-CIO v. Wisconsin Employment Relations Commission, 427 U.S. 132 (1976)). The essence of the *Machinists* preemption doctrine is that Congress intended some conduct that the NLRA neither protects nor prohibits to be left entirely unregulated. Employers sought to characterize *Machinists* preemption as doing something beyond leaving some conduct unregulated by the NLRA; employers argued that *Machinists* created a group of subjects (i.e., mandatory subjects of collective bargaining) that are free from *any* regulation.

[152] *Metropolitan Life Insurance Co.*, 471 U.S. at 753.

[153] *Id.* at 755.

gested that to hold that federal law in effect prohibited states from setting minimal employment standards would cause federal preemption to effect deregulation: it would "artificially *create* a no-law area."[154]

The analysis that the Court employed in rejecting the NLRA preemption claim in *Metropolitan Life* could be equally applicable in the ERISA context, had the Court not already concluded that the language and legislative history of the ERISA preemption provision prove unambiguously that preemption must be extremely broad. The Court demonstrated in its discussion of NLRA preemption that it entertained concerns about the deregulatory consequences of federal preemption. Further, since the Court rejected the argument that Congress intended the NLRA to leave certain terms of employment to the free market, it could have rejected the analogous argument about ERISA. Yet the Court declined to draw the connection, instead adhering to the contractualist vision of employee benefits that it had flirted with in *Alessi* but had spurned in the second part of *Metropolitan Life*.

Not until several years later, when the Court again confronted the absurdity of the distinction between insured and uninsured plans in *FMC Corporation v. Holliday*,[155] did any Justice publicly acknowledge that the Court's interpretive choice had perhaps been a mistake. However, by then the Court apparently regarded itself as committed to the line of reasoning it had already taken. In *FMC Corporation*, the Court reviewed a provision in Pennsylvania's Motor Vehicle Financial Responsibility Law that prohibited "any program, group contract or other arrangement" from seeking subrogation or reimbursement from a participant's tort recovery in a motor vehicle accident case.[156] The Court held that, although the Pennsylvania law was an insurance law that would be saved by the insurance savings provision, it could not be applied to a self-insured ERISA health plan such as the one maintained by FMC because of the "deemer clause," which prohibits ERISA plans from being deemed to be insurance. As a consequence, the law could not be applied to

---

[154] *Id.* at 757 (quoting Taggart v. Weinacker's, Inc., 397 U.S. 223, 228 (1970) (concurring opinion) (emphasis in *Taggart*)). This was the criticism of broad *Machinists* preemption that commentators had previously made. *See, e.g.*, Archibald Cox, *Recent Developments in Federal Labor Law Preemption*, 41 OHIO STATE L.J. 277 (1980).

[155] 498 U.S. 52 (1990).

[156] *Id.* at 55.

self-funded plans, although it could be applied to plans that purchase insurance.[157]

Returning to the plain language analysis of *Shaw*, the Court began by noting that the Pennsylvania law "has a 'reference' to" ERISA plans because it specifically mentions "'benefits payable by a hospital plan corporation.'"[158] The Court further determined that the state law "also has a 'connection' to ERISA benefit plans" because it "requires plan providers to calculate benefit levels in Pennsylvania based on expected liability conditions that differ from those in States that have not enacted similar antisubrogation legislation."[159] The Court then rejected two constructions of the deemer clause urged by the employee and an amicus, either of which would have narrowed the scope of preemption. On those readings, self-insured ERISA plans would be exempt only from state laws that "apply to insurance as a business, such as laws relating to licensing and capitalization," or state insurance laws that are "pretexts for impinging upon core ERISA concerns,"[160] which was the position the court of appeals had taken in the case.[161] These interpretations of the deemer clause would have narrowed the difference between insured and self-insured plans for purposes of ERISA and would have enlarged the range of state laws that could survive preemption. The Court rejected these constructions as being "unsupported by ERISA's language."[162]

As Justice Stevens pointed out in dissent, there is no reason for treating self-insured plans differently from insured plans, or for denying to beneficiaries of the former the state law protections that are available to beneficiaries of the latter. If Congress had so intended, Stevens reasoned, it would have said so, and in any event the entire mess could be avoided by a narrower reading of either the preemption clause or the deemer clause.[163] In Stevens' view, because the legislative history of section 514 showed that Congress was primarily concerned with overlap between federal and state requirements for plans, the "relates to" language ultimately adopted by Congress is "best explained as

---

[157] *Id.* at 61.
[158] *Id.* at 58–59. The Court quoted and relied on *Shaw's* plain language analysis.
[159] *Id.* at 59.
[160] *Id.* at 56.
[161] FMC Corporation v. Holliday, 885 F.2d 79, 86 (3d Cir. 1989).
[162] *FMC Corporation*, 498 U.S. at 63.
[163] *Id.* at 66.

an editorial amalgam of the two bills rather than as a major expansion of the section's coverage," as the Court evidently believed.[164] Turning to the deemer clause, Stevens pointed out that it was probably motivated by concern that states would subject ERISA plans to the same detailed licensing and capitalization requirements that apply to insurance companies.[165] Finally, Stevens pointed out that the Pennsylvania law is an example of the many state laws that apply to insurance companies as well as others and that regulate the business of insurance "but do not require one to be an insurance company in order to be subject to their terms."[166] Thus, he concluded, there was no reason in the language, history, or purpose of ERISA, nor any reason of policy, to preempt the application of such laws to self-insured plans.

The debate between the majority and Justice Stevens over the purpose of ERISA turned on whether the desirability of national uniformity of regulation for self-insured plans should take precedence over the desirability of allowing equivalent state law protections for beneficiaries of insured and self-insured plans. Not surprisingly, the statute itself yields no clear answers. In the majority's view, the principal goal of preemption is national uniformity, which will simplify plan administration for large employers and, indirectly, benefit beneficiaries of such plans.[167] The insurance exception to preemption is thus a necessary but undesirable accident of the tradition of regulating insurance at the state rather than federal level. In Stevens' view, the principal purpose of the statute was to protect plan participants, and from that perspective there is no reason to distinguish between insured and self-insured plans or to preempt state laws more broadly than necessary to avoid conflict between state and federal law. He thus saw no reason to construe the preemption provision broadly or the insurance savings provision narrowly. Nothing in the statute itself can definitively resolve the debate over which should be the preeminent purpose; the majority's insistence on its own reading of the language silenced what might have been

---

[164] *Id.* at 67.

[165] *Id.* at 69 (discussing a Missouri case, decided while ERISA was being considered, that subjected a pension plan to insurance licensing requirements, Missouri v. Monsanto Co., Cause No. 259,774 (St. Louis Cty. Cir. Ct., Jan. 4, 1973), *rev'd*, 517 S.W.2d 129 (Mo. 1974)).

[166] *Id.* at 70.

[167] *Id.* at 64–65.

a useful debate over how the Court should choose between these two plausible legislative purposes.

The distinction between insured and self-insured plans also produced significant unintended policy consequences. The majority approach created strong incentives for plans to self-insure in order to minimize exposure to regulation. To receive the benefits of some insurance protection, plans often purchase stop-loss insurance for claims above a certain amount. This raises the question whether states can regulate plans indirectly by regulating the stop-loss insurance, just as they could if the plan were fully insured. Courts have answered in the negative.[168] The widespread use of stop-loss insurance suggests that the Court's distinction between insured and self-insured plans is artificial, and the distinction has been criticized on this basis.[169] State insurance regulators fear that new forms of stop-loss insurance are really ordinary insurance with a high deductible and thus are essentially a subterfuge to evade state regulation.[170]

The development of the Court's ERISA preemption jurisprudence was not consistent. Whereas in some cases the Court appeared to be mainly textualist, at other times the Court combined an ostensible focus on language with a more significant examination of statutory purpose in an effort to reach sensible limits on preemption without abandoning its commitment to the notion that the language of the preemption provision can resolve cases. In two cases, the Court looked to the meaning of the term "employee benefit plan" to discern limitations on the scope of preemption. Since section 514 preempts state laws relating to "plans," the Court concluded that state laws that do not relate to plans, but only to "payroll practices" or "conditions of employment," are not preempted. The first of these cases was *Fort Halifax Packing Co. v. Coyne*,[171] holding that a Maine statute requiring severance payment in the event of a plant closing was not preempted by ERISA. Initially pursuing a plain language approach, the Court held that ERISA preempts laws relating to

---

[168] Thompson v. Talquin Bldg. Prods. Co., 928 F.2d 649 (4th Cir. 1993); Drexelbrook Engineering Co. v. Travelers Ins. Co., 710 F. Supp. 590 (E.D. Pa. 1989).

[169] *See, e.g.*, Jeffrey G. Lenhart, *ERISA Preemption: The Effect of Stop-Loss Insurance on Self-Insured Plans*, 14 VA. TAX REV. 615 (1995).

[170] HEALTH, EDUCATION, AND HUMAN SERVICES DIVISION, U.S. GAO, EMPLOYER-BASED HEALTH PLANS: ISSUES, TRENDS, AND CHALLENGES POSED BY ERISA, REP. NO. GAO/HEHS 95-167 (July 1995). Three states as well as the National Association of Insurance Commissioners have adopted laws intended to reduce this practice. *Id.*

[171] 482 U.S. 1 (1987).

employee benefit "plans," and that a statutory requirement that a one-time severance payment be made was not a "plan."[172] Turning from the language to the purpose of the statute, the Court reasoned that "pre-emption of the Maine statute would not further the purpose of ERISA pre-emption."[173] The point of ERISA preemption, according to the Court, was to permit employers with employee benefit plans to comply with a single set of administrative requirements regarding the payment of benefits. "A patchwork scheme of regulation would introduce considerable inefficiencies in benefit program operation, which might lead those employers with existing plans to reduce benefits, and those without such plans to refrain from adopting them."[174] Finally, the Court considered the consequences of preemption: it would make no sense for ERISA to preempt the Maine statute, because that statute "fails to implicate the regulatory concerns of ERISA itself . . . . The focus of [ERISA] is on the administrative integrity of benefit plans—which presumes that some type of administrative activity is taking place."[175] Because the Maine statute had nothing to do with an employee benefit plan, "[i]t would make no sense for pre-emption to clear the way for exclusive federal regulation, for there would be nothing to regulate."[176]

The Court, following the same analysis it had employed in *Metropolitan Life*, also rejected the employer's argument that the NLRA preempted the state law. The Court reasoned that the Maine statute merely set the backdrop against which the parties negotiated, just as state common law did. This is plainly correct.

---

[172] *Id.* at 8. "The Maine statute neither establishes, nor requires an employer to maintain, an employee benefit *plan*. The requirement of a one-time, lump-sum payment triggered by a single event requires no administrative scheme whatsoever to meet the employer's obligation." *Id.* at 12.

[173] *Id.*

[174] *Id.* at 11.

[175] *Id.* at 15.

[176] *Id.* at 16. The Court rejected the broad reasoning of the Maine high court, which had held that ERISA does not preempt state laws mandating the creation of benefit plans. Fort Halifax Packing Company, in a solicitude for employee protection that evidently did not extend to paying severance benefits in the event of plant closure, apparently had expressed concern that adherence to the rule adopted by the Maine Supreme Court would "create the opportunity for employers to circumvent ERISA's regulatory requirements by persuading a State to require the type of benefit plan that the employer otherwise would establish on its own," and that such a plan would presumably not be subject to any of ERISA's protections. *Id.* at 16. To avoid this result, the Supreme Court saved from preemption only state laws regarding employee benefits that have no effect on "plans," not the broader range of laws that the Maine court sought to protect.

In some respects, state law gives employers advantages: common law generally grants employers the right to run the workplace as they wish, absent a collective bargaining agreement restraining that right. And sometimes, as in this case, state law gives employees rights: here, the right to severance pay from employers.[177] To the extent that different state entitlements "complicate" negotiations for a nationwide collective bargaining agreement, the Court has not seen that as a serious impediment to the goal of a uniform national labor law. Yet the Court has never consistently held this view about ERISA preemption, as is clear from *Alessi*. The difference between *Fort Halifax* and *Metropolitan Life*, on the one hand, and *Alessi*, on the other, is the Court's unarticulated perception that the former cases involved the state mandating entitlements for all workers, while the latter concerned a state trying to interfere in contractual benefits relationships that were already established. The distinction is not analytically sound, but it is one that has never been exposed or defended.

In another case challenging a state law mandating benefits for all workers, the Court candidly acknowledged that the language of section 514 is unhelpful and proceeded quickly to focus on the purpose of ERISA. *Massachusetts v. Morash*[178] involved a state law mandating payment of unused vacation benefits to a terminated employee. The Commonwealth of Massachusetts had instituted a criminal proceeding against the president of a state bank for failing to pay discharged employees their full wages, including unused vacation time. The bank officer argued that ERISA preempted the Massachusetts statute because the bank's vacation policy was an employee benefit plan.[179] The Court began by noting that forty-seven states, the District of Columbia, and the United States all had similar wage payment laws and that over half included vacation pay as did the Massachusetts statute.[180] Although the Court commenced its preemption inquiry,

---

[177] Justice White, joined by Justices Rehnquist, O'Connor, and Scalia, dissented. In their view, "[a] state law 'which requires employers to pay employees specific benefits' clearly relate[s] to 'benefit plans' as contemplated by ERISA's pre-emption provision." *Id.* at 24 (quoting *Shaw*, 463 U.S. at 97). In the dissent's view, the Court's rationale created a loophole that would "allow States to effectively dictate a wide array of employee benefits that must be provided by employers" by simply characterizing them as not requiring the creation of an administrative scheme. *Id.* at 23.

[178] 490 U.S. 107 (1989).

[179] *Id.* at 108–09.

[180] *Id.* at 109–10.

just as it had in *Fort Halifax*, by trying to find the answer in the statutory definition of what is and is not an employee benefit plan, the Court quickly noted that ERISA's definition of "employee benefit plan" is circular: an "employee benefit plan" is defined as a "plan."[181]

Finding a plain language analysis unhelpful, the Court looked to the purpose of ERISA. Once the Court did so, a sensible answer to the preemption problem began to seem clear. If an employer does not maintain a separate fund for payment of benefits, a policy of paying benefits may not be a plan, because the purpose of ERISA is to prevent "the mismanagement of funds accumulated to finance employee benefits and the failure to pay employees benefits from accumulated funds."[182] According to the Court, "[b]ecause ordinary vacation payments are typically fixed, due at known times, and do not depend on contingencies outside the employee's control, they present none of the risks that ERISA is intended to address."[183] When vacation pay schemes require the creation of a fund, such as a multi-employer fund involving workers "who regularly shift their jobs from one employer to another," the protective concerns of ERISA are implicated, and the preemption provision would apply.[184]

What the Court found most convincing, however, were the undesirable consequences of preemption. Preemption of state laws such as the Massachusetts statute would "displace the extensive state regulation of the vesting, funding, and participation rights of vacation benefits; because ERISA's vesting and funding requirements do not apply to welfare benefit plans, employees would actually receive less protection if ERISA were applied to ordinary vacation wages paid from the employer's general assets."[185] The Court's effort in *Morash* to link the scope of preemption to the scope of protection provided by ERISA was unique until *New York Blues*.

Yet, reading *Morash* alongside *FMC Corporation*, it is easy to see the indeterminacy of statutory "purpose." This indeterminacy makes reliance on statutory purpose problematic for the Court. In *Morash*, the Court characterized the purpose of the statute as the protection of workers through the regulation of benefit plans,

---

[181] *Id.* at 113.
[182] *Id.* at 115.
[183] *Id.*
[184] *Id.* at 120.
[185] *Id.* at 119 (citations omitted).

and therefore suggested that state protective legislation only tangentially related to employee benefits should survive.[186] In *FMC Corporation*, the Court viewed ERISA as protecting workers by facilitating the formation and administration of plans, and thus minimizing the effect of state regulation;[187] in this analysis, preemption of state protective labor laws is consistent with ERISA's protective purposes. While the *FMC Corporation* view has dominated, *New York Blues* perhaps signals a resurgence of the *Morash* view.

The Court repeatedly asserted that broad preemption serves a fundamental purpose of ERISA, in that it encourages growth of the private employee benefit system by sparing plans and employer plan sponsors from the supposed inefficiencies that might result if plans were subject to state regulation.[188] The Court used a syllogism to articulate statutory purpose: the statute was intended to protect employees; benefit growth spawned by efficient management in the employee benefit system will be beneficial to employees in the long run; therefore, employees will benefit if plan sponsors are free of state regulation. The Court was evidently convinced that if plan sponsors find it unduly difficult to maintain plans, or if the law requires that plans be too generous to employees, plan sponsors will decide not to create plans or will reduce benefits. In *Shaw*, for example, the Court mused about how prohibiting the application of state antidiscrimination laws would in fact benefit employees rather than harm them:

> Obligating the employer to satisfy the varied and perhaps conflicting requirements of particular state fair employment laws, as well as the requirements of Title VII, would make administration of a uniform nationwide plan more difficult. The employer might choose to offer a number of plans, each tailored to the laws of particular States; the inefficiency of such a system presumably would be paid for by lowering benefit levels . . . . To offset the additional expenses, the employer presumably would reduce wages or eliminate those benefits not required by any State.[189]

---

[186] *Id.* at 115.

[187] FMC Corporation v. Holliday, 498 U.S. 52, 60 (1990) ("To require plan providers to design their programs in an environment of differing state regulations would complicate the administration of nationwide plans, producing inefficiencies that employers might offset with decreased benefits").

[188] *Id. See also* Fort Halifax Packing Co. v. Coyne, 482 U.S. 1, 10 (1987).

[189] 463 U.S. at 105 n.25 (1983).

The Court thus justified deregulation of employee benefits with speculation about economic behavior: the absence of state regulation ensures national uniformity, national uniformity ensures efficiency, and efficiency protects employees.[190] This is reminiscent of the Court's *Lochner*-era solicitude for the "right" of employees to contract for substandard working conditions.[191]

3. Chaos in the Lower Courts

The Court's textualism produced chaos in the lower courts.[192] Its emphasis on the meanings of the "relates to" clause, the "insurance savings" clause, and the "deemer" clause did not provide guidance to the courts with the primary responsibility for deciding thousands of ERISA preemption cases. Many lower courts designed preemption tests which differed markedly from those of the Supreme Court. The Ninth and Tenth Circuits, for example, developed a four-part test that had little relation to the Court's decisions.[193] The First Circuit held that workers' compensation laws affecting all employers are not preempted, even if the laws affect ERISA plans offered by some employers.[194]

---

[190] In suggesting that national uniformity is desirable or was intended by Congress, the Court had to confront the problem that ERISA does not explicitly displace other federal law and that many federal laws rely on states to set standards or to enforce federal mandates. The Court attempted to reconcile its view of preemption with the federal law savings provision as applied to Title VII, which itself relies on state law, by speculating the following:

> Congress might well have believed, had it considered the precise issue before us, that ERISA plans should be subject only to the nondiscrimination provisions of Title VII, and not also to state laws prohibiting other forms of discrimination. By establishing benefit plan regulation "as exclusively a federal concern," Congress minimized the need for interstate employers to administer their plans differently in each State in which they have employees.

Id. at 105 (quoting Alessi v. Raybestos-Manhattan, Inc., 451 U.S. 504, 523 (1981)).

[191] *See generally* Fisk, *Lochner Redux*, *supra* note *.

[192] ERISA preemption is not the only area of law where the Court's reliance on textualism created confusion for lower courts. One scholar has traced a similar phenomenon in cases under 42 U.S.C. § 1981 in the wake of the Court's textualist decision in Patterson v. McLean Credit Union, 491 U.S. 164 (1989). *See* George H. Taylor, *Textualism at Work*, 44 DEPAUL L. REV. 259 (1995).

[193] Under this test, ERISA preempts state laws: (1) regulating terms of plans; (2) creating reporting, disclosure, funding, or vesting requirements; (3) calculating the amount of benefits to be paid by plans; or (4) providing remedies for actions arising out of the administration of plans. Martori Bros. v. James-Massengale, 781 F.2d 1349 (9th Cir. 1986) (holding that ERISA does not preempt calculation of make-whole award based on fringe benefits); Airparts Co. v. Custom Benefit Servs., 28 F.3d 1062 (10th Cir. 1994) (holding state law claims of negligence, implied indemnity, and fraud against expert benefit plan consultant not preempted by ERISA).

[194] Combined Management, Inc. v. Superintendent of the Bureau of Ins., 22 F.3d 1, 3 (1st Cir. 1994), *cert. denied*, 115 S. Ct. 350 (1994).

The Second Circuit adopted a rule that a law indirectly affecting the cost of ERISA benefits is preempted if it operates in the "realm where ERISA plans must operate."[195] The Third Circuit rejected that rule and instead identified three factors to be used in determining whether ERISA preempts a state law.[196] One district court flatly rejected the Supreme Court's statement in *Greater Washington* that ERISA preempts any state law that specifically mentions or refers to ERISA plans, since following the Court's rule would have invalidated Michigan's method of taxing corporations based in part on labor costs, which of course include the costs of providing ERISA-covered benefits.[197] The Third Circuit also held that the Supreme Court's express reference rule does not apply when the express reference to the ERISA plan can be excised without changing the legal effect of the statute.[198]

Following the Supreme Court's language and cues about broad preemption, the lower courts found that ERISA preempted a wide variety of legislation having nothing to do with ERISA's purposes and concerns. There are far too many examples of the extraordinary breadth of ERISA preemption in the lower courts to note them all here.[199] One of the most egregious is presented by a series of cases in which the courts of appeals concluded

---

[195]NYSA-ILA Medical & Clinical Servs. Fund v. Axelrod, 27 F.3d 823 (2d Cir. 1994), *vacated and remanded sub nom.* Chassin v. NYSA-ILA Medical & Clincal Servs. Fund, 115 S. Ct. 1819 (1995).

[196]The three factors were:

> (1) whether the state law represents a traditional exercise of state authority; (2) whether the state law affects relations among the principal ERISA entities . . . rather than relations between one of these entities and an outside party, or between two outside parties . . .; and (3) whether the effect of the state law upon the ERISA plan is direct or merely incidental.

Travitz v. Northeast Dep't ILGWU Health & Welfare Fund, 13 F.3d 704, 709–10 (3d Cir. 1994).

[197]Thiokol Corp. v. Roberts, 858 F. Supp. 674 (W.D. Mich. 1994).

[198]United Wire, Metal & Mach. Health & Welfare Fund v. Morristown Memorial Hosp., 995 F.2d 1179, 1192 (3d Cir. 1993), *cert. denied*, 114 S. Ct. 382 (1993).

[199]For example, some courts have held as preempted state taxes which tax plan transactions or affect a plan's assets or investments. *See, e.g.*, E-Systems, Inc. v. Pogue, 929 F.2d 1100 (5th Cir. 1991) (holding that ERISA preempts Texas's "administrative services tax"); Morgan Guaranty Trust Co. v. Tax Appeals Trib. of N.Y. State Dep't of Taxation & Fin., 599 N.E.2d 656 (N.Y. 1992) (holding that ERISA preempts New York's capital gains tax as applied to plan assets); *but see* Retirement Fund Trust of the Plumbers Indus. v. Franchise Tax Bd., 909 F.2d 1266 (9th Cir. 1990) (holding California tax levy not preempted); Firestone Tire & Rubber Co. v. Neusser, 810 F.2d 550 (6th Cir. 1987) (holding municipal income tax not preempted). Courts have held state laws apportioning liability for tort damages to be preempted when an ERISA plan provdided one of the possible sources of insurance. *See, e.g.*, Travitz v. Northeast Dep't ILGWU Health & Welfare Fund, 13 F.3d 704, 709–10 (3d Cir. 1994) (holding survivor statute preempted but wrongful death statute not preempted); Auto Owners Ins. Co. v. Thorn Apple Valley, Inc., 31 F.3d 371 (6th Cir. 1994) (holding no-fault insurance law

that ERISA preempts state prevailing wage laws and state regulation of the terms of employment of apprentices. While the Supreme Court's textualist approach seems to compel such preemption (state laws regulating wages or working conditions of apprentices "relate to" ERISA-covered apprenticeship training funds), Congress clearly did not intend to eliminate the long tradition of state regulation of the wages and working conditions of apprentices. Preemption is simply an inadvertent consequence of the traditional way that apprenticeship programs are structured in the construction trades.[200]

In *Boise Cascade Corp. v. Peterson*,[201] the Eighth Circuit held that ERISA preempts state regulation of the conditions of employment of apprentices in the construction trades. ERISA covers "apprenticeship or other training programs,"[202] and thus state laws which "relate to" employee benefit plans covered by ERISA are preempted. Although the court noted that the state had regulated high-pressure pipefitting since 1937 because it is "a very dangerous activity" where shoddy work can cause explosions,[203] and that the suit arose out of the state's effort to end growing disregard of the apprentice-to-journeymen ratios that had prevailed since the 1940s, the court nevertheless invalidated the regulation because it related to an employee benefit plan covered by ERISA.[204] The court buttressed its plain language analysis by referring to ERISA's purpose of ensuring national uniformity in all matters pertaining to employee benefits and by asserting that enforcement of state rules regulating apprentices would expose employers to "conflicting or inconsistent state and local regulations."[205] The court did not explain how these sorts of regulations differ from any other state occupational safety or employment

---

preempted); *but see* Winstead v. Indiana Ins. Co., 855 F.2d 430 (7th Cir. 1988) (holding no-fault insurance statute not preempted), *cert. denied*, 488 U.S. 1030 (1989).

[200] In addition to the cases discussed in the text, there are others reaching similar results on similar reasoning. *See, e.g.*, National Elevator Indus. v. Calhoon, 957 F.2d 1555, 1562 (10th Cir. 1992) (invalidating Oklahoma's prevailing wage law as applied to apprentices); General Elec. Co. v. New York State Dep't of Labor, 891 F.2d 25 (2d Cir. 1989) (holding New York's prevailing wage law preempted), *cert. denied*, 496 U.S. 912 (1990); Keystone Chapter, Assoc. Builders & Contractors, Inc. v. Foley, 37 F.3d 945 (3d Cir. 1994) (holding order of the Pennsylvania Prevailing Wage Appeals Board preempted).

[201] 939 F.2d 632 (8th Cir. 1991), *cert. denied*, 502 U.S. 1027 (1992).

[202] 29 U.S.C. § 1002(1)(A) (1988).

[203] 939 F.2d at 634.

[204] *Id.* at 638.

[205] *Id.* at 637.

laws, or if these laws would also be preempted as applied to apprenticeship programs.

The Ninth Circuit reached the same result in *Local Union 598, Plumbers and Pipefitters Industry Journeymen and Apprentices Training Fund v. J.A. Jones Construction Co.*,[206] holding that the California prevailing-wage statute for apprentices on public works created a funding obligation for a plan and was thus preempted.[207] The prevailing-wage law, the court held, interfered with the employment contract by "'fundamentally and directly alter[ing] the employer's negotiated obligations,'" and "'add[ing] an additional statutory requirement—the cost of which [was] to be borne by the employer—to a private employee benefit plan.'"[208] In response to the plan's argument that preemption of the state law left no regulation in its place, the court noted that ERISA preemption cleared the way for future congressional action on the issue of apprenticeship wages: "'section [514(a)] has cleared the decks for such provisions, should Congress choose to address this concern in the future.'"[209] This is a new theory of preemption: prospective preemption in anticipation of hypothetical future federal legislation.[210]

---

[206] 846 F.2d 1213 (9th Cir. 1988), *aff'd mem.*, 488 U.S. 881 (1988).

[207] *Id.* at 1219. The court found support for its position in the Second Circuit's decision in Stone & Webster Eng'g Corp. v. Ilsley, 518 F. Supp. 1297 (D. Conn. 1981), *aff'd*, 690 F.2d 323 (2d Cir. 1982), *aff'd mem.*, 463 U.S. 1220 (1983), which had invalidated a Connecticut statute requiring an employer to continue the health benefit coverage of an employee who was receiving workers' compensation benefits.

[208] *Id.* at 1219 (quoting 690 F.2d at 329).

[209] *Id.* at 1220 (quoting 518 F. Supp. at 1301).

[210] The Ninth Circuit had previously reached the same conclusion in a line of cases beginning with Bechtel Construction, Inc. v. United Brotherhood of Carpenters & Joiners, 812 F.2d 1220 (9th Cir. 1987), in which the court held that the NLRA preempted a state law setting the wages to be paid to apprentices. In *Bechtel*, the court reasoned that since the NLRA protects the collective bargaining process, and wages are a mandatory subject of collective bargaining, the NLRA prevents states from dictating the outcome of wage negotiations by regulating wages. *Id.* at 1225. Recognizing that Metropolitan Life Insurance Co. v. Massachusetts, 471 U.S. 724 (1985), rejected the argument that the NLRA preempts state mandated minimum benefits, the Ninth Circuit determined that *Metropolitan Life* saved only *minimum* labor standards from preemption. Because the court of appeals interpreted the California law as permitting the parties to negotiate a wage lower than the state-set wage, the California law was not a minimum standard and was therefore not saved from preemption. *Bechtel*, 812 F.2d at 1222, 1225–26. Consequently, the *Bechtel* decision grants employers of unionized employees a power that non-union employers lack—the ability to pay apprentice wages below the level set by state law. The flaw in the court's analysis is obvious: "There is not the slightest reason to suppose that Congress intended to allow unions and employers, acting jointly, to establish employment conditions that a state forbids employers to establish unilaterally or by individual bargain." Cox, *supra* note 27, at 297. The court justified the anomalous treatment of unionized employees by assigning a "supreme value" to the collective bargaining process that trumps the

The Ninth Circuit expanded its rationale to invalidate California regulation of the training and pay for apprentices on public works projects in *Hydrostorage, Inc. v. Northern California Boilermakers Local Joint Apprenticeship Committee*.[211] In that case, an administrative agency had ordered Hydrostorage to comply with the state law, based on a finding that Hydrostorage had willfully failed to adhere to various state law requirements regarding the number of apprentices on the job site and had failed to make contributions to the training fund as required by statute.[212] Hydrostorage dashed into federal district court and obtained relief from the administrative order.[213] The Ninth Circuit determined that the fund was an employee benefit plan covered by ERISA and that the administrative order was preempted by ERISA because it was a state law relating to a covered plan.[214]

These cases illustrate that the Supreme Court's twin rationales for its ERISA preemption decisions—plain language and the importance of freeing plan sponsors (employers) from state regulation—were not only unhelpful to lower courts, but misleading and pernicious as well. The Court failed to develop a doctrine to guide the lower courts, and the language that it did provide invited results that did violence to any plausible congressional intent or statutory purpose.

---

regulation that could be applied to nonunion negotiations. To allow state regulation of working conditions for apprentices "would subordinate the bargaining process for all tradespeople (not just apprentices) to the goal of establishing uniform apprenticeship wages at all job sites." *Bechtel*, 812 F.2d at 1224. The court concluded that this "subjugation of the collective bargaining principle" was an unreasonable construction of the California statute, because to allow a state agency to set wages "would in effect give apprentices more than one representative, in violation of fundamental principles of federal labor law." *Id.*

[211] 891 F.2d 719 (9th Cir. 1989), *cert. denied*, 498 U.S. 822 (1990).

[212] *Id.* at 729.

[213] *Id.* at 730.

[214] *Id.* at 732. The court also rejected the fund's contention that the state regulation of apprentices was saved from ERISA preemption by § 514(d), the provision of ERISA that saves other federal law from preemption. The fund had urged that the state regulation was pursuant to the Fitzgerald Act, which directs the Secretary of Labor to formulate labor standards to protect apprentices and "to cooperate with State agencies engaged in the formulation and promotion of standards of apprenticeship," 29 U.S.C. § 50, and that the state regulation was saved by ERISA's savings of other federal law. The Ninth Circuit was unimpressed: "Assuming [the state law] was adopted in furtherance of the objectives of the Fitzgerald Act, it clearly is not an enforcement mechanism of federal law and to the extent orders under this section are preempted by ERISA, federal law is not impaired." 891 F.2d at 731.

B. *Why The Court Resorted to Textualism*

As illustrated in the preceding discussion, the Supreme Court's commitment to textualism has generated confusing law, outcomes that are inconsistent with the statutory purpose of employee protection, huge incursions on state regulatory authority inconsistent with the Court's supposed respect for the authority of state governments,[215] and a great deal of litigation. Given the serious consequences of these decisions, both for workers and for advocates of federalism, and given the Court's infrequent agreement regarding federalism and the balance of power between workers and firms, the predominance of unanimous ERISA opinions is surprising.[216] Furthermore, because Justices Blackmun, Brennan, and Marshall authored preemption opinions that significantly constricted the enforceability of state protective laws, the Court's concerns must have been more complex than simple anti-employee bias.

In this section, I offer two related explanations for why the Supreme Court adhered to textualism. One is that textualism was the reason for the decisions, and the other is that it was a rationale for decisions made, at least in part, on other grounds.[217] As to the first explanation, there are institutional reasons for relying on the plain meaning of language to decide cases. I will argue that the Court used textualism because the statute lacks—and the Court sought to avoid developing—a coherent vision of regulatory federalism that is an essential premise of an intelligible ERISA preemption analysis. The Court may have resorted to textualism in part because it offered an intellectually respectable basis on which to decide cases in an area of law that some or all members of the Court did not care to understand.

The second theory explaining the Court's decisions posits that the Court had a substantive vision that made the results the Court reached through textualism seem obvious, or at least plausible. The Court may have believed that ERISA established employee benefits as an exclusively federal concern, similar to the

---

[215] *See, e.g.*, Rice v. Santa Fe Elevator Corp., 331 U.S. 218, 230 (1947).

[216] Of the Court's 14 ERISA preemption decisions, 10 were unanimous, 1 had one dissenting vote, 1 had 3 Justices dissenting, and 2 were decided 5-4.

[217] I thus distinguish between two processes in judicial decisionmaking, that of deciding the case and that of providing a justification for the decision. Richard A. Wasserstrom, who identified the difference, calls the former "the process of discovery" and the latter "the process of justification." RICHARD A. WASSERSTROM, THE JUDICIAL DECISION 27 (1961).

vision that animated the Court's jurisprudence under the National Labor Relations Act. According to this theory, the Court was committed to national legislative dominance until it recognized that the result was being transformed into a substantive vision of laissez faire. As the extreme laissez faire implications of broad preemption became apparent, the Court began to struggle with preemption and with the textualist methods it had until then employed.

1. Textualism as a Reason

The "plain meaning" approach to statutory interpretation, according to Professor Frederick Schauer, functions less as a basis for accurately interpreting legislative intent or meaning than as a way of enabling judges to reach a decision about that intent or meaning.[218] Writing opinions that rely on the plain meaning of the statute facilitates development of a position upon which a majority of the Justices will agree. Judges on a multi-member court therefore might use plain meaning as a decisional rule simply to ease decisionmaking in those cases in which the judges have little knowledge, interest, or concern for the outcomes.[219] In such cases, Schauer explains, "where the substance of the dispute seems to the Justices . . . less politically or morally or economically charged, . . . jurisprudential views about methods of legal decisionmaking . . . are more likely to dominate."[220]

Some of the Supreme Court's ERISA preemption cases illustrate the phenomenon that Schauer has identified. *District of Columbia v. Greater Washington Board of Trade*,[221] *Shaw v. Delta Air Lines, Inc.*,[222] and, to a lesser extent, *FMC Corporation v. Holliday*,[223] *Ingersoll-Rand v. McClendon*,[224] and *Fort Halifax Packing Co. v. Coyne*,[225] focused on the meaning of the language of section 514. Although only *Greater Washington* purported to

---

[218] Frederick Schauer, *Statutory Interpretation and the Coordinating Function of Plain Meaning*, 1990 SUP. CT. REV. 231; *see also* Frederick Schauer, *The Practice and Problems of Plain Meaning: A Response to Aleinikoff and Shaw*, 45 VAND. L. REV. 715 (1992).

[219] Schauer, *Statutory Interpretation*, *supra* note 218, at 254; Schauer, *The Practice and Problems of Plain Meaning*, *supra* note 218, at 723.

[220] Schauer, *Statutory Interpretation*, *supra* note 218, at 248.

[221] 113 S. Ct. 580 (1992).

[222] 463 U.S. 85 (1983).

[223] 498 U.S. 52 (1990).

[224] 498 U.S. 133 (1990).

[225] 482 U.S. 1 (1987).

rely exclusively on plain meaning, several of the opinions illustrate the coordinating function of strict textualism. The Court relied on the language and, in some cases, the legislative history, which rejected narrower preemption and emphasized uniformity of regulation. The Court seemed oblivious to the nuances of employee benefits and unaware that language in its opinions would drastically alter the enforceability of state laws far removed from ERISA's concerns. This cannot be explained simply by a desire to immunize employers from regulation, since the Court's first two ERISA preemption opinions, *Alessi v. Raybestos-Manhattan, Inc.*[226] and *Shaw*,[227] were unanimous opinions written by Justice Marshall and Justice Blackmun, respectively. These decisions dramatically reduced the scope of operation for state protective laws, even in areas where ERISA provided inadequate protection, thereby producing far-reaching and undesirable consequences for employees and participants in benefit plans. These results seem flatly inconsistent with Justice Marshall's and Justice Blackmun's usual solicitude for workers' rights.[228] Additionally, Justice O'Connor authored *Pilot Life Ins. Co. v. Dedaux*,[229] another unanimous decision, in which the Court held that insurance bad faith tort claims were preempted; this view that state law should have no role seems inconsistent with Justice O'Connor's clear preference for federalism.[230]

Conversely, the Court also decided by unanimous opinion two cases during the 1980s that one might have expected to produce dissents from conservative members of the Court. In *Metropolitan Life Ins. Co. v. Massachusetts*[231] and *Massachusetts v. Morash*,[232] the Court used reasoning that elevated ERISA's protective purposes above the importance of national uniformity to hold that ERISA did not preempt state insurance and wage payment statutes. Both the reasoning and the result of these two decisions are difficult to reconcile with the Court's other ERISA preemption cases, and they are more protective of employees

---

[226] 451 U.S. 504 (1981) (holding a workers' compensation anti-offset provision invalid).

[227] 463 U.S. 85 (1983) (holding pregnancy discrimination provisions preempted).

[228] *See, e.g.*, Wards Cove Packing Co. v. Antonio, 490 U.S. 642, 661 (1989) (Blackmun, J., dissenting, joined by Marshall, J.); Patterson v. McLean Credit Union, 491 U.S. 164, 188 (1989) (dissenting opinion of Brennan, J., joined by Blackmun, J., and Marshall, J.).

[229] 481 U.S. 41 (1987).

[230] *See, e.g.*, New York v. United States, 505 U.S. 144 (1992).

[231] 471 U.S. 724 (1985).

[232] 490 U.S. 107 (1989).

than one would expect from some of the Justices. The predominance of unanimous opinions in deciding what have proven to be important questions of employment law is puzzling, because one would not expect such an ideologically divided Court to be in complete accord.[233]

Textualism can explain how the Court could have achieved unanimity in many decisions that had significant, ideologically charged consequences. On the surface, these decisions seemed to involve relatively unimportant and technical questions of the meaning of ERISA's preemption provision, insurance savings clause, and deemer clause. Focusing at this level obscured the significance these decisions would have in the struggle between employers and insurers on the one hand and workers and consumer groups on the other. If the political, social, and economic ramifications of the decisions had been clear to the Justices (as, for example, the social significance of the technical burden of proof issues in employment discrimination litigation are apparent to them[234]), there likely would have been fewer unanimous opinions.[235]

---

[233] I am not the first to see in the predominance of unanimous ERISA opinions a lack of attention to the detail or the significance of the law. Professor Langbein similarly criticized the Court's decision in Firestone Tire & Rubber Co. v. Bruch, 489 U.S. 101 (1989):

> The Supreme Court's opinion in *Bruch* garbles long-settled principles of trust law, [and] confuses trust and contract rubrics . . . . *Bruch* is such a crude piece of work that one may well question whether it had the full attention of the Court. I do not believe that either Justice O'Connor or her colleagues who joined this unanimous opinion [footnote omitted] would have uttered such doctrinal hash if they had been seriously engaged in the enterprise.
>
> Unfortunately, *Bruch* is not the first instance in which the Supreme Court has discharged ERISA business shoddily. [Langbein here cites as examples Connolly v. PBGC, 475 U.S. 211 (1986); Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724 (1985); PBGC v. Gray, 467 U.S. 717 (1984); and International Brotherhood Teamsters v. Daniel, 439 U.S. 551 (1979).] I understand why a Court wrestling with the grandest issues of public law may feel that its mission is distant from ERISA . . . . If the court is bored with the detail of supervising complex bodies of statutory law, thought should be given to having that job done by a court that would take it seriously.

John H. Langbein, *The Supreme Court Flunks Trusts*, 1990 SUP. CT. REV. 207, at 228–29.

[234] The Court's protracted and divisive struggle over burdens of proof in employment discrimination cases manifests underlying disagreement about the existence or pervasiveness of bias in employment. *See, e.g.*, St. Mary's Honor Center v. Hicks, 113 S. Ct. 2742 (1993); *see generally* Deborah C. Malamud, *The Last Minuet: Disparate Treatment After* Hicks, 93 MICH. L. REV. 2229 (1995).

[235] Interestingly, in two cases in which the Court did not rely principally on plain meaning in its decisions, but instead relied on the protective purpose of ERISA to find state laws saved from preemption, the Court split 5-4. *See Fort Halifax Packing Co.*, 482 U.S. 1 (1987); Mackey v. Lanier Collection Agency & Serv., Inc., 486 U.S. 825 (1988).

2. Textualism as a Rationale

The Schauer theory is not, however, a complete explanation. Textualism does not appear to have been the sole reason for the decision in some ERISA preemption cases. In many cases, it was not exclusively invoked, but was instead one reason among others that the Court gave for reaching a result. In addition to the textualist approach, the Court may have operated on a set of assumptions about federal dominance in regulating labor that made its interpretation of the language seem obvious.

The Court may have concluded that ERISA preempted state laws of all kinds because Congress appeared to have nationalized the entire employee benefits relationship, just as Congress had nationalized the entire union-management relationship under the NLRA. As long as the Court had confidence in the adequacy of the federal regulation, the scope of preemption seemed relatively straightforward, even though cases at the margins would always be difficult. When confidence in the federal scheme—and in Congress's ability to maintain that scheme's coherence—began to erode, so too did the Court's confidence in the exclusivity of federal law. ERISA preemption became problematic in part because the relationship between state and federal law in the whole field of labor and employment law became problematic.

The broad preemption of state legislation that the Court chose for ERISA was consistent with one of the Court's approaches to preemption under the National Labor Relations Act.[236] This is true even though ERISA preemption is express and NLRA preemption is implied.[237] The Court has read the NLRA to preempt not only state laws that conflict with specific provisions of the NLRA[238] or with the power of the National Labor Relations Board to define what constitutes permissible and prohibited ac-

---

[236] *See supra* text accompanying notes 143–149 (discussing NLRA preemption holding in Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724 (1985)).

[237] The NLRA is largely silent on the relationship between federal and state law. It mentions the role of state law in only two instances: Section 14(b) of the Taft-Hartley Act grants states the option of adopting "right to work" statutes, and § 14(c)(2) allows states to assert jurisdiction over labor disputes as to which the NLRB has declined to assert jurisdiction. 29 U.S.C. §§ 164(b), 164(c)(2) (1988).

[238] *See, e.g.*, ARCHIBALD COX ET AL., CASES AND MATERIALS ON LABOR LAW 987 (11th ed. 1991) (citing Bethlehem Steel Co. v. New York State Labor Rel. Bd., 330 U.S. 767 (1947); La Crosse Tel. Corp. v. Wisconsin Employment Rel. Bd., 336 U.S. 18 (1949); Plankington Packing Co. v. Wisconsin Employment Rel. Bd., 338 U.S. 953 (1950)).

tivity,[239] but also laws that conflict with what the Court believes Congress intended to be left unregulated by state or federal law.[240] The Court's statements that the NLRA broadly preempts state law in the field of labor relations were premised on the view that the NLRA and affiliated statutes[241] comprehensively regulate the relations between workers, unions, and employers, and that state courts would be hostile to the aims and methods of the federal scheme. Textualism therefore was not the sole basis on which the Court found meaning in section 514; textualism may also have been a jurisprudentially palatable rationale for decisions that seemed intuitively obvious based on the Court's belief in the dominance of federal labor law. The Court may have assumed that ERISA broadly preempted state law because ERISA seemed to fit within the established tradition of dominant and exclusive federal regulation of labor.

The Court's reliance on textualism became problematic when the significance of ERISA preemption and the ideological agenda of employers in arguing preemption became apparent to the Court and to commentators.[242] Justice Stevens was the first member of the Court to perceive the consequences of ERISA preemption for employees, and he was the first to dissent from the textualist approach, pointing out that in some cases the Court's view of the plain meaning of "relate to" led to absurd or unjust results. For instance, Justice Stevens stated in dissent in *FMC Corporation v. Holliday*[243] that the majority's analysis made little sense "[f]rom the standpoint of the beneficiaries of ERISA plans—who after all are the primary beneficiaries of the entire statutory program . . . ."[244] Similarly, when Justice Thomas wrote for the Court in *Greater Washington*[245] and relied on the same dictionary definition of "relates to" that the Court had used without dissent in *Shaw*, Justice Stevens dissented. Justice Stevens protested that the "growing emphasis on the meaning of the words

---

[239] San Diego Bldg. Trades Council v. Garmon, 359 U.S. 236 (1959).

[240] Lodge 76, Int'l Ass'n of Machinists v. Wisconsin Employment Relations Comm'n, 427 U.S. 132 (1976).

[241] *See, e.g.*, The Labor Management Reporting and Disclosure Act of 1959, 29 U.S.C. §§ 401–531 (1988).

[242] *Cf.* O'Neil, *Protecting ERISA Health Care Claimants*, *supra* note 27, at 723–24 (noting that ERISA "now often serves as a shield for employers, insurance companies, and plan administrators, rather than to protect participants' rights").

[243] 498 U.S. 356 (1990).

[244] *Id.* at 366 (Stevens, J., dissenting).

[245] 113 S. Ct. 580 (1992).

'relate to'"[246] had diminished reliance on common sense in ERISA preemption cases, and he urged the Court "to take a fresh look at the intended scope of the preemption provision that Congress enacted."[247] Perhaps Justice Stevens' view finally garnered the support of the entire Court last Term in *New York Blues* because it was a case in which the consequences of following the Court's prior plain language reasoning would have been both especially preposterous (invalidating almost all state regulation of hospital charges) and far afield from the Court's vision about national dominance in labor relations.

ERISA preemption has been vexing just as labor law preemption has been vexing for the Court. In both areas, the Court has decided a disproportionately large number of preemption cases but failed to bring clarity to the law.[248] Under both the NLRA and ERISA, the Court stated that the field was comprehensively regulated by federal law and that the areas about which the federal statute was silent were best left without regulation.[249] But neither statute is actually comprehensive. Labor preemption became problematic in part because the significance of collective

---

[246] *Id.* at 586 n.6 (Stevens, J., dissenting).

[247] *Id.* at 586 (Stevens, J., dissenting).

[248] *See, e.g.*, Cox, et al., *supra* note 238, at 959 ("No legal issue in the field of collective bargaining has been presented to the Supreme Court more frequently in the past thirty years than that of the preemption of state law, and perhaps no other legal issue has been left in quite as much confusion."). The Court's most recent attempt to reconcile its NLRA preemption cases was Livadas v. Bradshaw, 114 S. Ct. 2068 (1994), in which it held that the NLRA preempted a state labor commissioner's policy of declining to enforce a wage payment statute against employers of unionized employees. The Court emphasized that state law forms the backdrop against which parties negotiate collective bargaining agreements, and that it defeats the goal of the NLRA to deprive unionized employees of the protections of state law. The Court used this same reasoning to reject NLRA preemption in Metropolitan Life Ins. Co. v. Massachusetts, 471 U.S. 724, 752–58 (1985), but declined to use that same reasoning in ERISA cases. *See supra* text accompanying notes 151–154.

[249] *Compare* Lodge 76, Int'l Ass'n of Machinists v. Wisconsin Employment Relations Comm'n, 427 U.S. 132, 149 (1976); Archibald Cox, *Labor Law Preemption Revisited*, 85 HARV. L. REV. 1337, 1352 (1972); *and* Shaw v. Delta Air Lines, Inc., 463 U.S. 85, 98–99 (1983).

The evolution in Archibald Cox's thinking on labor preemption tracked and shaped the rise and fall of enthusiasm about broad preemption. Cox initially favored broad federal preemption, *see, e.g.*, Archibald Cox, *Federalism in the Law of Labor Relations*, 67 HARV. L. REV. 1297 (1954), but he later had some second thoughts, at least as to particular topics. Cox, *supra* note 27. Cox's doubts about very broad preemption apparently influenced the Court's decision to retreat from the unnecessarily broad language of *Lodge 76*. *See Metropolitan Life*, 471 U.S. at 753 (citing Cox); New York Telephone Co. v. New York State Dept. of Labor, 440 U.S. 519 (1979) (citing Cox in several places and holding that NLRA does not preempt New York unemployment insurance law which pays benefits to strikers).

bargaining as a source of protection for workers has declined in relative importance compared to state law,[250] just as ERISA preemption has become problematic because of the increased significance of state regulation of health benefits. The Court quickly realized that its suggestion in *Machinists* that the holes in the NLRA are intended to create space for laissez faire was an overstatement, and it retreated from the overbroad *Machinists* preemption.[251] The Court did not have the same flexibility about ERISA preemption because of its early commitment to a textualist approach to section 514. The Court has not yet decided how much of the area not directly regulated by federal law under ERISA is intended for laissez faire and how much remains a proper subject for state regulation. Congress itself did not provide an answer. As Congress avoids this task, either through silence or through drafting ill-considered language, the burden of determining the appropriate relationship between federal and state law in employee benefit regulation shifts to the judiciary.[252] It is to that task that I now turn.

III. The Need for Pragmatism and for a Theory of Regulatory Federalism

Because ERISA does not definitively resolve the proper balance between state and federal law in matters of employee benefits, courts should approach the preemption inquiry pragmatically. The question should *not* be the meaning of "relates to," but rather whether allowing employers to be subject to state regulation would defeat the goals of protecting employee expectations in receiving benefits. The ERISA preemption inquiry requires an appraisal of the need for national uniformity balanced against

---

[250] *See* Katherine Van Wezel Stone, *The Legacy of Industrial Pluralism: The Tension Between Individual Employment Rights and the New Deal Collective Bargaining System*, 59 U. Chi. L. Rev. 575 (1992); Gottesman, *supra* note 27, at 361–62; Silverstein, *supra* note 27, at 28–33.

[251] *Metropolitan Life*, 471 U.S. at 754–57.

[252] It is not uncommon for Congress to remain silent on important matters when enacting legislation:

> The hard fact of political life is that, in order to draft a bill that can pass both Houses of Congress and garner a presidential signature, it is sometimes politic to leave some things unsaid. But that political decision is also a judgment to delegate those matters to the courts without much direction.

Ronald D. Rotunda, *The Civil Rights Act of 1991: A Brief Introductory Analysis of the Congressional Response to Judicial Interpretation*, 68 Notre Dame L. Rev. 923, 927 (1993).

the advantages of regulation at the state level. This calls for pragmatism.

A pragmatic approach to statutory interpretation requires courts to recognize the inevitability of statutory ambiguity, and further requires a self-conscious process of deciding cases in accordance with the goals of the legislation. The judge's role should complement the legislative process, for the legislative process is inherently only a part of the lawmaking process. The approach I advocate is something like that described most prominently by William Eskridge, Jr., Philip Frickey, Daniel Farber, and Cass Sunstein.[253] As applied to ERISA preemption, a "pragmatic" judge would develop a theory of regulatory federalism to assess whether ERISA should supersede a particular state law.

The preemption questions that scholars and courts *should* consider differ from the preemption questions that judges normally consider. This different concept of regulatory federalism is not even a distant cousin of the general notion of federalism that judges reflexively invoke to create a presumption against preemption.[254] First, it is not a constitutional argument. There is no constitutionally compelled reason for courts to revise preemption doctrine. My argument is entirely functional: courts should modify ERISA preemption doctrine because the current doctrine makes no policy sense and is not dictated by the statute. Second, in suggesting that courts should consider regulatory federalism in deciding ERISA preemption cases, I am not advocating that courts necessarily strive to save state law from preemption, which is what the usual federalist presumption is supposed to do. A presumption in favor of state law would not necessarily result in a more coherent ERISA doctrine or in greater loyalty to the

---

[253] This is a drastic oversimplification, of course, and it asks the reader to set aside for the moment the question of the legitimacy of the judicial role thus defined. *See, e.g.*, William N. Eskridge, Jr., *Dynamic Statutory Interpretation*, 135 U. PA. L. REV. 1479 (1987); William N. Eskridge, Jr. & Philip P. Frickey, *Statutory Interpretation as Practical Reasoning*, 42 STAN. L. REV. 321 (1990); Daniel A. Farber, *The Inevitability of Practical Reason: Statutes, Formalism, and the Rule of Law*, 45 VAND. L. REV. 533 (1992); T. Alexander Aleinikoff, *Updating Statutory Interpretation*, 87 MICH. L. REV. 20 (1988); Sunstein, *supra* note 19, at 160–233. Influential earlier versions of the same notion were articulated by GUIDO CALABRESI, A COMMON LAW FOR THE AGE OF STATUTE (1982), and HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW (William N. Eskridge, Jr. & Philip P. Frickey, eds., 1994).

I have borrowed the term "complement" from Jane S. Schacter, *Metademocracy: The Changing Structure of Legitimacy in Statutory Interpretation*, 108 HARV. L. REV. 593, 626–36 (1995).

[254] *E.g.*, Cipollone v. Liggett Group, Inc., 505 U.S. 504, 518, 532–33, 542 (1992).

legislative goals. In short, whether the traditional presumption in favor of federalism is one of statutory interpretation, as the Court suggested in *Gregory v. Ashcroft*,[255] or constitutional law, as the Court suggested in *New York v. United States*,[256] federalism is not a reason to decide a case; it is a rationalization. Usually federalism is asserted as a rationale to support decisions that seem to be motivated more by views on, for example, civil rights than on state sovereignty as a value in itself.[257]

If, however, the courts were to take seriously their assertion that federalism principles command a presumption that Congress did not intend to supersede state laws in areas of traditional state concern, I submit that ERISA preemption of some state regulation of welfare benefit plans violates their oath of fealty to state power. Notwithstanding the existence of an express preemption provision, congressional intent to invalidate the vast range of state laws that have fallen prey to ERISA preemption is anything but "manifest." Thus, the Court's ERISA preemption cases are inconsistent with its assertion that Congress cannot, through the existence of its preemption power, create a "federally mandated free market" unless its intent to do so is "clear and manifest."[258]

What the courts ought to ask themselves in deciding ERISA preemption cases, therefore, is a pragmatic question: to what extent will decentralization of regulatory authority over this area of law facilitate or hamper the sensible operation of the law? If this were the question, then courts could pay attention to something that ought to be relevant in assessing ERISA preemption of state law—the fact that Congress paid different degrees of

---

[255] 501 U.S. 452 (1991).

[256] 505 U.S. 144 (1992).

[257] My views on the nature and values of federalism are in accord with those of critics of federalism. *See, e.g.*, Edward L. Rubin & Malcolm Feeley, *Federalism: Some Notes on a National Neurosis*, 41 UCLA L. REV. 903 (1994). These authors persuasively argue that federalism does not do any of the things normally claimed to its credit. As they put it, federalism (as opposed to decentralization) "does not secure citizen participation, does not make government more responsive or efficient by creating competition, and does not encourage experimentation." Nor, they say, does it diffuse governmental power or secure community. *Id.* at 909. *See also* Erwin Chemerinsky, *The Values of Federalism*, FLA. L. REV. (forthcoming 1995); Erwin Chemerinsky, *Rehabilitating Federalism*, 92 MICH. L. REV. 1333 (1994) (book review).

[258] Puerto Rico Department of Consumer Affairs v. ISLA Petroleum Corp., 485 U.S. 495, 500 (1988). In the *ISLA Petroleum* case, the Court considered the possibility that federal regulation of some aspects of a contractual relationship could operate through broad preemption to create a federally mandated free market. The Court held that Congress could deregulate petroleum allocation and pricing by preempting state law, but that Congress's intent to use preemption to create a free market regime must be clear. The Court failed to find the requisite clarity of intent. *Id.*

attention to different subject matters of ERISA. For example, Congress focused far more on the problem of adequate funding of pension plans than on the restrictions that should be placed on a plan's or an employer's power to eliminate or modify coverage in a health benefit plan. Moreover, Congress paid no attention in ERISA to the regulation of the terms of apprenticeship programs. These facts about the coverage of ERISA ought to be relevant when courts decide whether to invalidate state law.

The Court took a step in the right direction in *New York Blues*.[259] The Court stated, "[w]e simply must go beyond the unhelpful text and the frustrating difficulty of defining its key term, and look instead to the objectives of the ERISA statute as a guide to the scope of the state law that Congress understood would survive."[260] The Court then asserted that the objective behind section 514(a) was to "minimize the administrative and financial burden of complying with conflicting directives . . . ."[261] After reconciling its prior cases under this new pragmatic approach, the Court looked at the purpose and effect of the state law to decide whether it imposed unacceptable administrative or financial burdens. Accepting that the state law would affect the cost of providing benefits in New York, the Court concluded that this indirect economic effect "does not bind plan administrators to any particular choice"[262] about which benefits to provide, and does not "preclude uniform administrative practice or the provision of a uniform interstate benefit package if a plan wishes to provide one."[263] Therefore, the Court believed that there was no conflict with ERISA, because "cost-uniformity [between states] was almost certainly not an object of pre-emption, just as laws with only an indirect economic effect on the relative costs of various health insurance packages in a given State are a far cry from those conflicting directives from which Congress meant to insulate ERISA plans."[264]

The Court's assumptions that the purpose of ERISA preemption was to minimize administrative and financial burdens on

---

[259] 115 S. Ct. 1671.

[260] *Id.* at 1677.

[261] *Id.* (quoting Ingersoll-Rand Co. v. McClendon, 498 U.S. 133, 142 (1990)). Although I disagree with this characterization of the objective of § 514(a), I do agree that the Court used the appropriate method of interpretation in looking at the objective of the statute to define the scope of preemption.

[262] *Id.* at 1679.

[263] *Id.*

[264] *Id.* at 1680 (internal quotation marks omitted).

interstate plans, and that broad preemption serves that goal, are open to question. First, the legislative history does not suggest that Congress intended preemption to minimize administrative or financial burdens on plans irrespective of the harm that elimination of state law could cause to plan participants. Second, invalidation of state regulation may not always serve the statute's protective purposes. As long as the provision of income security is a cost of doing business, national uniformity may be desirable as a way of eliminating competition and creating economies of scale for large employers subject to uniform national requirements. But whether, and the extent to which, either of these propositions is true is an empirical question. For example, it has been argued that environmental regulation at the federal level may not be necessary to prevent states from competing for industry by offering pollution standards that are too lax.[265] I do not know whether national uniformity is more efficient for business or will avoid a "race to the bottom" in the environmental field, the employee benefits field, or any other area of federal law. The point is that these are not by themselves arguments for broad preemption, although they are usually offered as such, without some empirical basis for assessing whether they are valid assertions.

One needs to be similarly concrete about the desirability of state regulation, for state regulation may be consistent with ERISA's purposes and requirements. ERISA was enacted against a backdrop of extensive state regulation of employment, health care, and insurance. Although Congress has authority to legislate on these subjects, it is not obvious that ERISA should be read to effect a broader displacement than is necessary to foster ERISA's objectives. To shift from mostly state to mostly federal control of these areas would cause confusion during transition and would add to the workload of a Congress that already has too little time to keep all the existing federal statutes up to date. Indeed, decentralized regulatory authority may foster ERISA's protective purposes. For instance, it may be (although this too is an empirical question) that state legislatures are more responsive to the concerns of consumer and worker groups than is Congress, because it is less expensive to mount a successful

---

[265] Richard L. Revesz, *Rehabilitating Interstate Competition: Rethinking the "Race-to-the-Bottom" Rationale for Federal Environmental Regulation*, 67 N.Y.U. L. REV. 1210 (1992).

lobbying campaign in most states. Thus, although significant health care reform may be blocked at the federal level by a well-organized and well-financed business and insurance lobby, it may not be blocked at the state level. Alternatively, the facts may be otherwise: it may that the states do not make good "laboratories."[266] These various possibilities illustrate that a pragmatic approach to interpreting ERISA's preemption provision should avoid the tendency to rely on unexamined and perhaps erroneous assumptions such as "federalism," "states as laboratories," and "national uniformity equals regulatory efficiency."

Having explained what a pragmatic approach to ERISA preemption is, let me be clear about when and why such an approach is appropriate. I do not believe that a pragmatic approach should be used to interpret every provision in every statute. The pragmatic approach is appropriate only when, as is the case with some aspects of ERISA preemption, the statutory language is ambiguous and there are otherwise no clear answers from the structure or history of the legislation. This will be particularly true when Congress first regulates in a field, as with ERISA, or when it undertakes major revision of existing legislation.

Textual theories of interpretation make major reform legislation terribly problematic,[267] especially when the new federal law is intended to displace any substantial amount of state law. Congress dramatically changed the landscape of employee benefits and had many big problems and small details to consider. Therefore, it is not surprising that it failed to define precisely which state laws should survive. Thus, pragmatism is necessary when, as in the case of ERISA, Congress obviously fails to consider fully the effect of a new federal law on a complex body of state regulation.

But the failure of congressional oversight is not the only justification for pragmatism about ERISA preemption. If the impossibility of congressional oversight were the only justification for courts to adopt a pragmatic approach, I would have to address the obvious argument that Congress could have, but con-

---

[266] "It is one of the happy incidents of the federal system that a single courageous State may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country." New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting). *See* Laguarda, *supra* note 24, at 191 (arguing that the states have not been good laboratories for health care reform).

[267] SUNSTEIN, *supra* note 19, at 113–22.

sciously did not, adopt different preemption standards for state laws relating to pensions and fiduciary behavior (which ERISA extensively regulates) than for welfare benefits (which ERISA does not comprehensively regulate). The dispute about enforcement of state legal ethics rules against legal services plans, it could be argued, alerted Congress to the hazards of broadly preempting state laws only indirectly related to ERISA's requirements, and Congress forged ahead with broad preemption anyway. Thus, the argument would go, judicial fine-tuning cannot be justified as a remedy for legislative oversight. However, ERISA is not simply a case of oversight.

Even if Congress had had time to consider the implications of section 514's language carefully, it could not have formulated *ex ante* a policy that would decide all cases. It may be that when Congress passes a statute covering a relatively narrow subject, it can decide in advance all the preemption issues that are likely to arise and resolve them itself. But when it enacts a lengthy and complex statute that displaces state law and regulates across the scope of the employment relationship (as well as financial relationships far removed from employment), Congress simply cannot anticipate all the preemption problems that are likely to arise.[268]

Unless very detailed, statutory language is often necessarily more of a statement of principle than a specific directive. Proceeding by the common law method is therefore inevitable, as the Supreme Court itself has noted. For example, in interpreting the preemption provision of the Airline Deregulation Act of 1978, which uses the same "relates to" language as does ERISA,[269] the Court struggled with the same problems of overbroad language and recognized the need for case-by-case accommodation between state and federal law.[270]

Moreover, when Congress regulated as broadly as it did in ERISA, it could not have anticipated how context or social change would fundamentally alter the significance of the preemption of state laws pertaining to health benefits. For example, extremely broad preemption has had consequences for health

---

[268] Preemption is often (though not always) a less difficult problem when the federal statute makes federal law a floor, not a ceiling, for state authority. *See* California Fed. Sav. & Loan Ass'n. v. Guerra, 479 U.S. 272 (1987).

[269] 49 U.S.C.A. § 1305(a)(1) (1988).

[270] *See* American Airlines, Inc. v. Wolens, 115 S. Ct. 817, 819 (1995); *see also* Morales v. Trans World Airlines, Inc., 504 U.S. 374 (1992).

care financing that Congress could not have anticipated in 1974. State laws like the one at issue in *New York Blues*[271] address problems in the health insurance market that were not perceived in 1974. Nor could Congress have considered the effect of ERISA on medical malpractice litigation against HMOs and insurance utilization review firms, because both were relatively unusual in 1974.

In any event, whether or not Congress could have resolved the issues about the desirable relationship of state and federal law in the ERISA context, Congress plainly did not do so in section 514. Therefore, courts, whether they admit it or not, are forced to do it. In two extreme cases, where the courts were blatantly wrong,[272] Congress corrected judicial interpretations. But Congress plainly does not have the time to oversee each of the hundreds of ERISA preemption issues that arise in litigation each year and to correct each of the failures. Courts ought not decide cases on the assumption that Congress will correct all of the errors they make, or that Congress's failure to change a statute necessarily constitutes an endorsement of the results.

There are two main arguments against pragmatism: that judges ought not, for reasons of legitimacy, decide cases based on their notions of enlightened public policy, and that judges cannot competently do so. Both arguments hold that if statutes are flawed, as ERISA's preemption provision clearly is, the onus is on Congress to fix them.[273] Even if the preemption provision is

---

[271] 115 S. Ct. 1671.

[272] At least, this was the case in the judgment of some in Congress who decided to expend their political capital on ERISA issues.

[273] Congressional efforts to "fix" problems with ERISA preemption have had a mixed history. On the one hand, Congress did amend the preemption provision in response to decisions holding that ERISA preempted state marital property laws that protected a non-employee spouse's interest in an employee spouse's pension. *See supra* note 61. Furthermore, Congress amended the preemption provision in response to a determination that ERISA preempted Hawaii's law requiring employers to provide health benefits for employees. *See infra* text accompanying notes 281–283. Also, Congress amended § 514 to clarify the role of state laws in regulating multiple employer welfare arrangements (MEWAs). 29 U.S.C. § 1144(b)(6) (1988).

But, as I explain below, Congress has failed to enact other bills that would save additional state laws from preemption. Notable examples include bills to overturn *Pilot Life*'s preemption of state insurance bad faith laws, *see* O'Neil, *supra* note 27, at 763–70 (describing myriad bills), to overturn court of appeal decisions regarding state regulation of apprenticeship programs, e.g., H.R. 1036, 103d Cong., 1st Sess. (1993), and to eliminate the different treatment of insured and self-insured plans, *see supra* notes 143–147 and accompanying text. While the failure to act may be taken as evidence that Congress approves of the outcomes in these cases, it may also mean that a well-organized group may prevent action to overturn an interpretation of the statute that would not be favored by Congress. *See infra* text accompanying notes 282–288.

the product of congressional error rather than a deliberate policy choice, so the argument goes, the courts should not fix the error. They should allow public excoriation to prod Congress into action. When courts attempt to fix congressional mistakes, they encourage congressional dereliction of duty. In this view, judges have the institutional role of forcing Congress to do its job better, not of taking on Congress's job themselves.[274]

The plain meaning theory may allay our doubts about judicial legitimacy.[275] The supposed virtue of plain meaning theory from a jurisprudential standpoint is the fiction that judges simply give life to the inherent meaning of the words, rather than choosing among several possible meanings, thus reducing judicial activism. The evidence, however, is to the contrary. Recent studies on the Supreme Court's new penchant for using dictionaries to decide cases suggest that the use of dictionaries does not constrain judicial activism. As Professor Pierce has shown, the Court has used plain language to overturn long-settled construction of statutes, to reject interpetations preferred by politically accountable administrative agencies, and to disregard clearly contrary legislative intent.[276] Thus, the dictionary has been a powerful weapon for a new brand of judicial activism. Moreover, the legitimacy of plain meaning theory rests on factual premises—that the legislature usually means what it says, and that it can express that meaning unambiguously—which are questionable in the case of ERISA's preemption provision. The intentionalist and purposive theories of interpretation find legitimacy in the view of judges

---

[274] A thoughtful statement of this position is found in Schacter, *supra* note 253, at 636–46. This was the view advanced by Bickel and Wellington in their article criticizing § 301 of the Labor Management Relations Act, 29 U.S.C. § 185 (1988), and by Justice Frankfurter in his dissent in Textile Workers Union v. Lincoln Mills, 353 U.S. 448 (1957). *See* Alexander M. Bickel & Harry H. Wellington, *Legislative Purpose and the Judicial Process: The Lincoln Mills Case*, 71 HARV. L. REV. 1, 26, 34 (1957). They argued that Congress defaulted on its responsibility to design a federal law governing collective bargaining agreements by creating federal jurisdiction to enforce collective agreements without at the same time enacting a body of rules for the federal courts to use in doing so.

[275] There are two jurisprudential defenses of textualism. In one view, judges are the agents and the legislature is the principal; the unelected judge simply carries out the legislative direction and does not pursue her own agenda. *See* SUNSTEIN, *supra* note 19, at 112–13. Alternatively, textualist theory may view the courts as "autonomous interpreters" who enforce the meaning of the statute as a reasonable person would understand it, not necessarily as the legislature intended it. In this view, democratic legitimacy of the unelected judge comes from him standing in the shoes of the person to whom the law applies and reading the law as that person would read it. *See* Merrill, *supra* note 132, at 353; Schacter, *supra* note 253, at 636–46.

[276] Pierce, *supra* note 21. *See also* Merrill, *supra* note 132.

as agents of the legislature, carrying out its predefined will. What I have said thus far about ERISA's history, however, suggests that current preemption doctrine bears little fealty to Congress's will, such as it was in 1974.

If one believes that courts should not correct statutory errors, one must believe either that Congress will fix the problem or that the cost of its failing to do so is worth the gain in judicial legitimacy. As for the former idea, it is doubtful that opponents of pragmatism really believe that Congress will fix the problems, although their theory forces them to pretend that they do. The comparatively small number of amendments to section 514 and the large number of problems that remain suggest that Congress will not fix the Court's errors. As for the latter idea, a great deal of unintentionally irrational law is not a fair price for a small fig leaf of judicial legitimacy.

ERISA is an excellent example of the classic observation that it is a great deal more difficult for Congress to correct flawed statutes than it is to enact them in the first place.[277] The reason is that interests coalesce around the advantageous aspects of the status quo. If legislative action is the only method of correcting statutory errors, then error will be the inevitable result of Congress's first stab at regulating in an area, unless Congress gets it entirely right the first time.[278] In the case of ERISA preemption, no one fully perceived in 1974 that broad preemption was a tremendous benefit for employers, insurers, and plans. However, these parties soon figured it out, and they have fought hard ever since to protect it.[279] The persistence of the unduly broad preemption language without amendment is thus an example of what Sunstein calls "statutory failure": the benefits of the language are significant for a highly organized though narrow group, while the costs may be great but are spread widely among a population that is not likely to organize effectively for change.[280]

---

[277] CALABRESI, *supra* note 253, at 6; GRANT GILMORE, THE AGES OF AMERICAN LAW 95 (1995).

[278] *Id.*

[279] *Compare* Schaffer & Fox, *supra* note 81, at 51 (arguing that although business firms later defended ERISA preemption of state regulation of health benefits on the ground that it interfered with freedom to design benefits packages, there is no evidence that they did so in 1974) *with* Robert R. Rosenblatt, *Health Reform: Tangled Up in a Knot of Deal-Killers*, L.A. TIMES, Aug. 21, 1994, at A1 ("America's biggest corporations will fight to the end any effort to let states write their own rules for regulating employee health benefits.").

[280] SUNSTEIN, *supra* note 19, at 102.

There is evidence in the history of efforts to amend the ERISA preemption provision to support this hypothesis. In 1974, just as Congress was putting the finishing touches on ERISA, the state of Hawaii enacted a statute that required employers to provide health insurance for their employees.[281] When ERISA took effect, business groups in Hawaii challenged the Hawaii Pre-Paid Health Care Act as being preempted by ERISA, and the courts so held.[282] Beginning in 1974, and for nine years thereafter, the senators from Hawaii waged a legislative battle to save their state statute from ERISA preemption. The Hawaiian senators' first bill simply eliminated all ERISA preemption of state health insurance laws; they encountered strong opposition and few allies. Later bills were less sweeping, seeking an exemption only for Hawaii's statute. Finally, in 1983, Congress amended the preemption provision to save the Hawaii statute. But the final version was not a complete victory for employees, as it saved only the 1974 version of the Hawaii statute, not the statute as it was modified by a 1976 amendment providing more generous benefits than the original version. Interestingly, the amendment to the ERISA preemption provision contained an unusual statement of legislative intent that made clear that only Hawaii's health care law was saved: "The amendment made by this section shall not be considered a precedent with respect to extending such amendment to any other State law."[283] Hawaii's success in saving the Health Care Act may be credited to the extraordinary legislative power of Senators Inouye and Matsunaga and to the perception by the rest of the Congress that Hawaii is in some sense unique. Moreover, the congressional correction hardly seems to be the result of careful consideration of the best way to accommodate state and federal law. Why, for example, would Congress require Hawaii to maintain the law enacted in 1974, rather than the later one, which may have been passed in response to changed conditions or experience? Bills that would allow Hawaii to make its system more generous to employees have been introduced in Congress, but have died there.[284]

---

[281] This story of the amendment to save the Hawaii statute is drawn from Schaffer & Fox, *supra* note 81, at 54–60. The story is repeated in Laguarda, *supra* note 24, at 179–85.

[282] Standard Oil Co. v. Agsalud, 633 F.2d 760 (9th Cir. 1980), *aff'd mem. sub nom.* Agsalud v. Standard Oil Co., 454 U.S. 801 (1981).

[283] Act of Jan. 14, 1983, Pub. L. No. 97-473, sec. 301(b), 96 Stat. 2605, 2612 (1983) (not codified).

[284] *E.g.*, S. 287, 103d Cong., 1st Sess. (1993); *see* 139 CONG. REC. S1174–75, (daily

A second example, one that Congress has yet to fix despite several years of legislative struggle, is drawn from the preemption of state laws regulating apprenticeship programs and the working conditions of apprentices. As described above,[285] courts of appeals held these state laws preempted, even though ERISA does not regulate the working conditions of apprentices and there is no evidence that Congress thought by enacting section 514 it was invalidating well-established state regulations regarding a whole category of workers.[286] Bills have been introduced to overturn these decisions, but none has been enacted.[287] Business, of course, strenuously objects to state regulation of the wages and working conditions of workers (including apprentices), and thus far apprentices and their allies have not mustered enough support to overcome determined business opposition. If Congress fails to amend ERISA in response to these literal, but mistaken, applications of the overbroad preemption provision, it allows unintended irrational consequences to persist. However, it is unclear how much would be gained as a practical matter by forcing Congress to spend its scarce time correcting judicial errors.

The question thus arises as to whether judges are competent to design a more sensitively calibrated preemption doctrine. The most obvious response is that judges have been doing precisely that for federal statutes which lack an explicit congressional statement regarding presumption. For all the criticism that various particular preemption doctrines have received,[288] there is no reason to believe that the task is necessarily beyond the ken of the federal judiciary. The real question is one of comparative institutional competence: are the courts or Congress better suited to make these sorts of determinations?[289] The competence ques-

---

ed. Feb. 3, 1993) (remarks of Sen. Akaka (D-Haw.)); S. 590, 102d Cong., 1st Sess. (1991); *see* 137 CONG. REC. S2932–33 (daily ed. Mar. 7, 1991) (remarks of Sen. Akaka). Senator Akaka reintroduced the bill in the 104th Congress, but it has not been enacted. S. 266, 104th Cong., 1st Sess. (1995); *see* 141 CONG. REC. S1443–45 (daily 3d. Jan.24, 1995) (remarks of Sen. Akaka).

[285] *See supra* notes 200–214 and accompanying text.

[286] *Id.*

[287] H.R. 1036, 103d Cong., 1st Sess. (1993); H.R. 2782, 102d Cong., 2d Sess. (1992). Both H.R. 1036 and H.R. 2782 were debated in the House, 139 CONG. REC. H8958–76 (daily ed. Nov. 9, 1993); 138 CONG. REC. H7274–96 (daily ed. Aug. 4, 1992), but neither became law.

[288] *See, e.g.*, Gardbaum, *supra* note 30 (discussing preemption generally); Wolfson, *supra* note 34; Drummonds, *supra* note 37 (discussing preemption in the labor and employment law context).

[289] *See generally* Neil K. Komesar, *A Job for the Judges: The Judiciary and the*

tion is answered in part by the Court's own recognition that case-by-case adjudication is sometimes necessary to accommodate state and federal law.[290] Even Congress seems to recognize that in certain circumstances some other entity may be better suited to make individualized assessments of the desirable scope of ERISA preemption. This recognition appears in a bill pending in Congress that would grant to an executive branch agency the authority to waive ERISA preemption for qualifying state health care reform plans at the same time it authorizes such plans under the Medicaid program.[291] As I have suggested, Congress simply cannot anticipate all preemption problems when it enacts a far-reaching law that displaces a substantial amount of state law. If there is no executive agency to work out an accommodation, the courts are the only institution capable of making the case-by-case assessment that is required.

IV. CONCLUSION

By last Term, judicial and scholarly discourse over the scope of ERISA preemption of state law had reached a dead end. While the lower courts continually complained that the language of the statute could not mean what it says, the Supreme Court persisted in pretending that the language had a meaning and that its meaning could decide cases. For years, scholars criticized the language of the statute and the results of the cases to no avail, until last Term in *New York Blues*. Now, the Court has finally admitted that the language of ERISA offers virtually no help to courts in deciding cases and has thus all but given the lower courts license to ignore it.

The Supreme Court's commitment to textualist interpretation was the main catalyst in the evolution of ERISA preemption from its origins as a last-minute compromise in a massive piece of new legislation to its current status as one of the most important aspects of health care and employment law policy. Judicial interpretation accorded section 514 significance in restricting state policy options in the fields of health care and employment

---

Constitution in a Massive and Complex Society, 86 MICH. L. REV. 657 (1988); Neil K. Komesar, *Taking Institutions Seriously: Introduction to a Strategy for Constitutional Analysis*, 51 U. CHI. L. REV. 366 (1984).

[290] *See* American Airlines, Inc. v. Wolens, 115 S. Ct. 817, 827 (1995).

[291] *See supra* note 1.

that Congress never intended. Ironically, the Court's textualism prevented the Court from fully appreciating the consequences of its decisions and made it difficult for the Court to change direction even when it became apparent how far ERISA preemption doctrine had strayed from the statute's protective purposes. The Court's recent change of emphasis evidenced in *New York Blues* indicates that only when the Court abandoned its reliance on textualism could it begin the difficult task of making sense of ERISA preemption. Thus, the story of the first twenty years of ERISA preemption doctrine is the story of the shortcomings of textualism as a method of statutory interpretation.

But this is more than just a case study of the failures of textualism. This history of ERISA preemption offers some valuable insights into general preemption doctrine as well. Foremost, it has taught us that the preemption doctrine's search for legislative intent is doomed to fail. Whether we admit it or not, courts are creating preemption doctrine in the ERISA context, as in many others, with little guidance from the legislature. As long as they are doing so, ERISA scholars should now do the research that will make the courts' jobs easier. It is time to think about and write about the proper balance between state and federal regulation of employee benefits rather than the meaning of section 514. While it is unwarranted hubris to suggest that this ought to be the last law review article written about the meaning of the words "relate to" in section 514 of ERISA, for the sake of us all, I hope it is.

ARTICLE

NEEDLE EXCHANGE, HIV TRANSMISSION, AND ILLEGAL DRUG USE: INFORMING LAW AND PUBLIC POLICY WITH SCIENCE AND RATIONAL DISCOURSE

STEVEN R. SALBU*

Efforts to control the spread of AIDS have been constrained by the controversial nature of many prevention programs. This is especially true of needle exchange programs. In this Article, Professor Salbu explores the issues behind needle exchange and the validity of arguments both for and against such programs. He also provides a comprehensive summary of research efforts to date and concludes that while research into the effectiveness of needle exchange programs is inconclusive, there is growing evidence that such programs reduce the spread of AIDS. These findings are, however, tentative, and the author proposes the conditional funding of future programs be contingent upon a primary research function and continued success.

Acquired Immune Deficiency Syndrome (AIDS) is among the greatest public health concerns in modern history.[1] It is caused by the human immunodeficiency virus (HIV), a blood-borne virus spread among humans via perinatal, blood-to-blood, or semen-to-blood contact.[2] Because HIV is blood-borne rather than

---

* Associate Professor, Graduate School of Business, University of Texas at Austin. B.A., Hofstra University; M.A., Dartmouth College; J.D., The College of William and Mary; M.A., Ph.D., The Wharton School of the University of Pennsylvania. The author of this Article may be reached via e-mail at SALBUS@CBACC.BUS.UTEXAS.EDU.

[1] The disease has generated an array of compelling social issues regarding prevention, treatment, resource allocation, and civil rights. While all these kinds of issues are vitally important, this Article focuses on one specific prevention issue. Commentators on health policy choices for the eradication of AIDS increasingly view prevention of new HIV transmissions as critical because the search for vaccinations and cures has proven so frustrating. *See, e.g.*, Norman Hearst, *AIDS Risk Assessment in Primary Care*, J. AM. BD. FAM. PRAC., Jan.-Feb. 1994, at 44 (noting physician's comment, "Prevention is our best weapon in the fight against the acquired immunodeficiency syndrome.").

[2] The author exercises caution in stating the means of HIV contagion, as much remains unknown about AIDS and the spread of HIV. Most scientific sources suggest that the disease is transmitted primarily or exclusively by these three methods. *See, e.g.*, Carolyn B. Britton, *HIV Infection*, 11 NEUROLOGIC COMPLICATIONS OF DRUG AND ALCOHOL ABUSE, NEUROLOGIC CLINICS 605, 605 (1993).

Pincus notes four methods of HIV transmission, recognized by the Centers for Disease Control (CDC) as the sole vectors:

> (1) Through sexual contact where there is an exchange of blood or semen (to or from any physical organ with access to the blood system); (2) By sharing intravenous drug needles; (3) Through blood transfusions or other nonsexual internal contact with contaminated blood or blood products, or (4) Through

airborne, opportunities for contagion tend to be concentrated among high-risk activities most likely to result in intermingling of blood, or of blood and semen. Among the high-risk practices associated with HIV transmission by the Centers for Disease Control (CDC) are male-to-male sexual intercourse; sexual relations with numerous partners, including male and female prostitutes; receipt of blood clotting products prior to 1985; and the injection of intravenous drugs.[3]

Early recognition of high-risk activities has informed the public health policy choices aimed at reducing the spread of AIDS. Because the practices considered to create the greatest risk of HIV transmission are voluntary behaviors,[4] they are subject to individual control. Accordingly, policymakers have developed AIDS programs that attempt to modify behavior, theoretically reducing the incidence of specific unsafe practices.[5] Perhaps in part because both sexuality (particularly homosexuality) and injection drug use are politically sensitive[6] and morally volatile[7] topics, policy recommendations aimed at slowing contagion rates

---

an infected mother to her child either during pregnancy or during post-partum breast feeding.

Laura B. Pincus, *HIV Education in the Legal Studies Curriculum*, 13 J. LEGAL STUD. EDUC. 75, 94 (1995). For further discussion of HIV transmission, see DEANNA E. GRIMES & RICHARD M. GRIMES, AIDS AND HIV INFECTION 30–33 (1994).

[3] Pincus, *supra* note 2, at 95. The notion of high-risk or higher-risk behaviors is socially and politically charged. The classification of certain behaviors as high-risk may be interpreted to imply that all other behaviors are low-risk. For example, one might infer from the specification of male-to-male anal intercourse that unprotected heterosexual intercourse is a lower-risk activity than it really is. While the behaviors listed here are certainly considered to be among the highest-risk activities, other behaviors may be associated with significant risk of HIV transmission as well.

[4] *See* text accompanying note 3.

[5] Safer sex campaigns and sexual abstinence campaigns are common programs aimed at reducing sexually transmitted HIV. Needle exchange programs and war on drugs programs have likewise been employed to try to reduce the transmission of the virus by needle sharing.

[6] Political pressures have frequently determined the nature of public policy decisions that are embodied in legislation, regulation, or the decision not to adopt legislative or regulatory recommendations. Peter Lurie of the University of California, San Francisco, has suggested that the primary obstacle to federal funding of needle exchange has been political rather than scientific. *Feds Hear Criticism of Stance on Needle Exchange*, ALCOHOLISM & DRUG ABUSE WK., Mar. 13, 1995, at 2.

[7] In one survey, 72% of respondents reported having little or no sympathy for people who contract AIDS through the use of illegal drugs. The study found that 60% of respondents reported the same sentiments regarding those who contract AIDS through homosexual activity. The figure fell dramatically to 20% in regard to all people with AIDS. *See* Robert J. Blendon et al., *Public Opinion and AIDS: Lessons for the Second Decade*, 267 JAMA 981, 983 (1992) (citing a CBA/*New York Times* poll taken by the Roper Center for Public Opinion Research in 1988). The divergence of attitudes based on source of HIV transmission reflects some societal disapproval of the behaviors of illegal drug users and homosexuals.

have tended to polarize along an abstinence-accommodation continuum. Proponents of abstinence-oriented measures promote public policies designed to eradicate behaviors that are risky, and which some view as inappropriate or immoral.[8] Proponents of accommodation-oriented measures de-emphasize or rebut inferences of moral turpitude in association with unsafe behavior.[9] They posit an approach that purportedly is more realistic than abstinence.[10] They contend that anal intercourse and injection drug use will continue, and that measures to slow the spread of AIDS should therefore focus on ways in which these behaviors can be made safer, rather than ways in which they can be expunged.[11] Among the accommodation-oriented policy approaches that have been widely considered and sporadically adopted are safer sex campaigns and needle exchange programs,[12] the latter of which are the subject of this Article.

The federal government has adopted neither needle exchange programs nor over-the-counter sales of syringes as part of a national AIDS policy.[13] Likewise, few state or municipal jurisdictions have authorized official needle exchange programs to date.[14] Yet the potential effectiveness of such programs has be-

---

[8] *Id. See also* Pedro Ruiz & Francisco Fernandez, *Human Immunodeficiency Virus and the Substance Abuser: Public Policy Considerations*, TEXAS MED., May 1994, at 64, 64 ("While an attitude of tolerance toward substance abuse prevailed during the 1970s in the United States, the prevailing attitude during the 1980s was one of intolerance. At present, many public demands call for the implementation of punitive measures against addicts.") (citation omitted).

[9] *See* Dawna Friesen, *Needle Exchange a Prickly Issue for St. Paul's*, VANCOUVER SUN, Mar. 3, 1995, at A1 (quoting the Chair of the Department of Family and Community Medicine at St. Paul's Hospital, "We took the position that the distribution of needles itself is not immoral.").

[10] *See, e.g.*, Richard Stevenson, *Harm Reduction, Rational Addiction, and the Optimal Prescribing of Illegal Drugs*, 12 CONTEMP. ECON. POL'Y 101 (1994) (observing that among harm reduction adherents, abstinence is seen as an unrealistic short-term goal for dependent drug users).

[11] The accommodation approach is typified by the Dutch model of illegal drug use control, typically labelled a "harm reduction" approach. For further discussion of the harm reduction approach, see *infra* notes 110–111, 400–405, and accompanying text.

[12] Needle exchange programs are also called "needle programs" in this Article. The terms "needle exchange" and "syringe exchange" are used interchangeably. For the purposes of public policy analysis, exchange of needle and exchange of syringe are very much analogous. The comments and recommendations in this Article refer to both.

[13] *See* Don C. Des Jarlais et al., *Continuity and Change Within an HIV Epidemic: Injecting Drug Users in New York City, 1984 Through 1992*, 271 JAMA 121, 126 (1994) ("Almost all developed countries except the United States have made legal access to sterile injection equipment—through syringe exchanges, over-the-counter sales, or both—a primary component of AIDS prevention for IDUs.") (citation omitted).

[14] Kurt L. Schmoke, *A Symposium on Drug Decriminalization: An Argument in Favor of Decriminalization*, 18 HOFSTRA L. REV. 501, 517 (1990).

come a critical public policy issue, since needle sharing is responsible for an ever-increasing percentage of HIV infections. By 1992, the CDC attributed twenty-three percent of all AIDS cases in the United States to injection drug use.[15] Particularly in urban areas of the United States,[16] injection drug use has begun to supplant high-risk sexual activities between men as the leading cause of HIV transmission.[17] Because the sharing of needles is a high-risk practice[18] responsible for a growing proportion of new AIDS cases,[19] governing bodies in many nations have considered the adoption of needle exchange programs.[20] Under these programs, a government agency or private organization typically provides free new needles to drug users in exchange for used needles.[21]

As we shall see in Part II, needle exchange programs are highly controversial.[22] Assessments of the wisdom of these programs range from vehement support to adamant condemnation. In response to the instinctive reactions of politicians, social com-

---

[15] MARY ELLEN HOMBS, AIDS CRISIS IN AMERICA: A REFERENCE HANDBOOK 9 (1992).

[16] For example, in Washington, D.C. during the first nine months of 1994, there were 226 new cases of AIDS among heterosexual drug users, as opposed to 199 new cases attributed to sexual activity by gay men. Amy Goldstein, *Most New Cases of AIDS in D.C. Hit Drug Users*, WASH. POST, Dec. 1, 1994, at A1.

[17] This phenomenon has been evident since the late 1980s and appears to be at least partially attributable to the success of educational efforts aimed at promoting safer sex in the gay community. By 1989, the proportion of all AIDS cases in the United States attributed to the use of injected drugs had grown to 27.5%. CDC ACQUIRED IMMUNODEFICIENCY SYNDROME (AIDS) MONTHLY SURVEILLANCE REPORT, Aug. 1989.

[18] Don C. Des Jarlais et al., *HIV Infection and Intravenous Drug Use: Critical Issues in Transmission Dynamics, Infection Outcomes, and Prevention*, 10 REV. INFECTIOUS DISEASES 151 (1988). The spread of AIDS through intravenous drug use includes transplacental or perinatal transmission from mothers who take drugs intravenously. Provisional Committee on Pediatric AIDS, American Academy of Pediatrics, 1992 to 1993, *Reducing the Risk of Human Immunodeficiency Virus Infection Associated with Illicit Drug Use*, 94 PEDIATRICS 945 (1994) [hereinafter *Illicit Drug Use Report*].

[19] *See, e.g.*, Goldstein, *supra* note 16. *See also* Marsha F. Goldsmith, *A Sticky Issue: HIV and the IVDU*, 266 JAMA 1053 (1991) ("Only 5% to 10% of the world's injecting drug users have thus far become infected, and yet needle sharing remains a common practice," creating the potential for "new explosive epidemics in communities of injecting drug users." (quoting statement of Jonathan Mann of the Harvard AIDS institute)).

[20] Holland, Great Britain, and Australia are among the nations that have developed and implemented various kinds of needle exchange programs. For discussion of some of these programs, see *infra* part III.

[21] For more detailed discussion of the specific characteristics of some needle exchanges, *see generally infra* part III.

[22] In a recent poll of 1001 randomly selected adults, 55% of respondents supported the use of needle exchange programs to stem the spread of diseases such as AIDS. *Drug Use Occupies Emergency Rooms, Public Opinion Polls*, 109 PUB. HEALTH REP. 586 (1994).

mentators, journalists, and others, scientists and social scientists have begun to study the effects, both positive and negative, of needle exchange.[23] To date, the data and findings from the studies are suggestive but inconclusive.[24] Moreover, the studies have been published in brief articles in science or social science journals, providing merely scattered bits of isolated information.[25] There is a compelling need for systematic synthesis of the various findings and analysis of their implications for public policy. This Article is intended to fill this void by providing both a thorough discussion of our knowledge concerning needle programs and a set of policy recommendations based on this knowledge.

Part I provides a discussion of the laws and regulations that affect the development and implementation of needle exchange programs. Part II briefly examines the public policy arguments for and against needle exchange. Part III contains a thorough discussion, critique, and analysis of scientific and social scientific literature on needle program effectiveness in reducing HIV transmissions. Part IV examines the nature and characteristics of needle programs that appear most likely to be effective. Part V discusses an area of investigation that has received relatively little treatment—the degree, if any, to which needle programs may increase the use of illegal drugs. Part VI contains recommendations concerning the legal status of needle exchange programs and the design of both the programs and the research projects that examine them.

I. Laws and Regulations that Affect the Development and Implementation of Needle Exchange Programs

Those who propose needle exchange as a means of curbing new HIV transmissions are impeded by legal and regulatory restraints that predate our experience with AIDS. Some states require a prescription for sellers to dispense needles and syringes, and for buyers to purchase them.[26] In a number of states, the distribution of hypodermic needles and syringes is illegal if

---

[23] *See infra* part III.

[24] *See infra* part III.

[25] *See infra* part III.

[26] *See, e.g.*, Cal. Bus. & Prof. Code § 4143 (West 1995) ("Except as otherwise provided by this article, no hypodermic needle or syringe shall be sold at retail except upon the prescription of a physician and surgeon, dentist, veterinarian, or podiatrist.").

the distributor knows or has reason to know that the apparatus will be used to inject drugs illegally.[27] A number of states also ban the possession of paraphernalia used for unlawful injection.[28]

These laws were adopted as part of the fight against illegal drug use, before the threat of AIDS rendered the strict control of needle and syringe possession arguably counterproductive. Restrictions on dispensing and receiving needles under paraphernalia control laws constrain the supply of new needles. This effect may inadvertently encourage needle sharing and the spread of HIV.[29] In response, a number of states have passed[30] or attempted to pass[31] laws expressly permitting the operation of needle exchanges. While the statutes that have been adopted to date expressly authorize the establishment of needle exchange programs,[32] proposed statutes in other states would create an exception to paraphernalia laws, under which needles and syringes would not be considered prohibited paraphernalia under specified conditions.[33]

Apart from the relatively straightforward statutory authorization of needle exchange programs in such states as Connecticut and Hawaii,[34] a handful of state courts and legislatures have addressed the issues that arise from needle exchanges that operate within the context of potentially hostile anti-drug laws. Likewise, the federal government has adopted a policy on the federal funding of local needle exchanges. These state and federal legal histories are provided in the subsections that follow.

---

[27] *See, e.g.*, CAL. HEALTH & SAFETY CODE § 11364.7 (West 1995).

[28] *See, e.g.*, CAL. HEALTH & SAFETY CODE § 11364 (West 1995).

[29] Lawrence Gostin, Panel on Needle Exchange and Bleach Distribution Programs, *Law and Policy, in* PROCEEDINGS: WORKSHOP ON NEEDLE EXCHANGE AND BLEACH DISTRIBUTION PROGRAMS, 113, 114 (Commission on Behavioral and Social Sciences and Education, National Research Council, and Institute of Medicine, sponsors, 1994).

[30] *See, e.g.*, CONN. GEN. STAT. § 19(a)-124, § 21(a)-65 (1978), as amended by 1990 Conn. Acts 90-214 (Reg. Sess.); D.C. CODE ANN. § 33-603.1 (1994); HAW. REV. STAT. § 325-112 (1994).

[31] For example, Governor Pete Wilson has successfully opposed several bills proposed in the state legislature that would have permitted needle exchanges in California. *Judge Blocks Needle Exchange*, UPI Regional News, Feb. 24, 1995, *available in* LEXIS, Nexis Library, UPI File.

[32] *See supra* note 30.

[33] *See Around Texas and Southwest*, DALLAS MORNING NEWS, Feb. 23, 1995, at 14A (noting this type of proposed legislation in Texas).

[34] *See supra* note 30.

A. *California*

In states like California, where paraphernalia control laws exist[35] but bills authorizing needle exchanges have not passed,[36] local governments as well as grassroots organizations have tried to circumvent the prohibitions of distribution and possession of drug paraphernalia in various ways. In several counties and municipalities, local authorities have instituted or sanctioned needle exchange programs that appear to be in contravention of California's paraphernalia control laws, justifying their actions by arguing that AIDS has created a public health emergency.[37] While the public health emergency approach to circumventing the paraphernalia control laws had some early success,[38] a lower level California court has recently held that a Sacramento needle exchange program violates the state's statute prohibiting the provision of hypodermic needles for illegal drug use.[39]

As in a number of California counties, Sacramento County supervisors had approved a needle exchange program as part of a local effort to fight AIDS. While Judge Ford recognized the supervisors' "laudable intentions of trying to prevent the scourge that is AIDS," his decision ultimately required that supervisors defer judgment to the longstanding state paraphenalia control

---

[35] *See supra* notes 26–28.

[36] *See supra* note 31.

[37] *See, e.g.*, Rick DelVecchio, *Needle Swap Supporters Make Pitch for Legal Status: Alameda County Official Declares an Emergency*, S.F. CHRON., Feb. 28, 1995, at A13 (noting declaration of a "public health emergency" by Alameda County, California interim public health officer Dr. Barbara Allen, to assist work by needle exchange activists); Clarence Johnson, *Boardwatch*, S.F. CHRON., Feb. 28, 1995, at A20 (reporting San Francisco's continued state of emergency in support of needle programs). One commentator has characterized the state of emergency approach as a means of getting around the law by "winks and nudges [that] make for plain stupid and inconsistent public policy . . . ." *Legalize Needle Exchanges*, S.F. CHRON., Mar. 10, 1995, at A22.

[38] In Alameda County, efforts by the District Attorney to prosecute needle distributors have been thwarted by jury acquittal. In one instance, the jury foreman suggested that the District Attorney should stop prosecuting needle distributors, and noted that jurors spent some time discussing whether the governor and legislature of California should legalize needle programs. The foreman, a retired police officer, reported that the jury agreed that the law had been broken but had been persuaded to accept a defense of necessity. *Needle Dispensers Acquitted*, SACRAMENTO BEE, Mar. 11, 1995, at B3.

[39] *California Judge Strikes Down Needle Exchange Program*, Reuters, Feb. 24, 1995, *available in* LEXIS, Nexis Library, Reuters File [hereinafter *Judge Strikes Program*]. The relevant statute provides, "Except as authorized by law, any person who delivers, furnishes, or transfers, possesses with intent to deliver, furnish, or transfer, or manufactures with intent to deliver, furnish, or transfer, drug paraphernalia, knowing, or under circumstances where one reasonably should know, that it will be used to . . . inject . . . or otherwise introduce into the human body a controlled substance . . . is guilty of a misdemeanor." CAL. HEALTH & SAFETY CODE § 11364.7 (West 1995).

statute.[40] Although the decision was only at the county superior court level, four out of five county supervisors stated that the county was unlikely to appeal.[41] The decision suggests that local efforts to circumvent state statutes, typically by declaring a state of emergency and attempting to exercise police powers in response to that emergency,[42] are likely to fail in California unless (1) the California legislature passes a statute expressly authorizing needle exchanges; (2) the California legislature passes a statute delegating to local authorities responsibility for determining policies to fight AIDS; or (3) state authorities decline to prosecute those local authorities who adopt needle programs.[43]

In the summer of 1995, the California Attorney General issued an opinion letter supporting the Sacramento County decision, stating that needle exchange programs violate California law notwithstanding either local declarations of public health emergency or the evocation of a defense of necessity.[44] The opinion letter asserts that the state legislature's ban on drug paraphernalia distribution was intended to pre-empt local regulation.[45] It notes that declaration of local emergencies can be utilized only to request resources and aid from other state political subdivisions, and cannot be employed to establish local authority that overrides state law in regard to issues of statewide concern.[46] The opinion letter also contends that the defense of necessity is limited in California and is not applicable to the promotion or facilitation of illegal conduct.[47] Steve Telliano, a spokesperson for the California State Attorney General, emphasized that the letter provided that official's interpretation of the law rather than a decision regarding prospective prosecutorial intent.[48] The Mayor

---

[40] Patrick Hoge, *Needle Exchange Illegal, Judge Rules*, FRESNO BEE, Feb. 26, 1995, at E31.

[41] *Id.*

[42] For example, Mayor Riordan's declaration of an AIDS emergency early in 1995 was viewed as an effort to facilitate needle exchange in Los Angeles in the face of the California paraphernalia statute. *Bottom Line: Needle Exchange Saves Lives: Pacoima Program Deserves Support, Not Moral Outrage*, L.A. TIMES, Feb. 19, 1995, at B18.

[43] *See Judge Strikes Program*, *supra* note 39 (citing statement by an attorney opposing the needle program that the ruling could be used to invalidate other programs in California).

[44] *See* Opinion of Attorney General Daniel E. Lungren, 94 Op. Att'y Gen. 1104 (1995).

[45] *Id.*

[46] *Id.*

[47] *Id.*

[48] Rachel Gordon, *Lungren: Needle Swaps are Illegal*, S.F. EXAM., June 17, 1995, at A1.

of San Francisco announced shortly thereafter that he would risk incarceration rather than cease operation of the city's needle exchange program.[49]

B. *Washington*

In *Spokane County Health District v. Brockett*,[50] the Supreme Court of Washington upheld the legality of the Spokane needle exchange despite provisions of Washington's version of the Uniform Controlled Substances Act (UCSA)[51] that forbid the distribution of drug paraphernalia under certain conditions.[52] The court found overriding statutory approval of needle exchange programs under the broad authority given to certain counties by the state's Omnibus AIDS Act.[53] The court held the Spokane County Health District's resolution to establish and implement a needle exchange to be a legitimate part of a Regional AIDS Network Plan developed under the Act.[54]

In essence, the court permitted local decision-making under the Omnibus AIDS Act to prevail over conflicting provisions of the state's Controlled Substances Act. The decision rested on the recognition of a broad, generic statutory authority permitting local health districts to protect the public health.[55] This authority was interpreted to require the liberal construction of public health statutes, the subject matter and expediency of which were held to be "'beyond judicial control,' except as they may violate some

---

[49] *Needle Swaps and Bad Law*, S.F. CHRON., June 20, 1995, at A16.

[50] 839 P.2d 324 (Wash. 1992).

[51] *See* WASH. REV. CODE § 69.50 (West 1995).

[52] *Brockett*, 839 P.2d at 327 ("It is unlawful for any person to deliver . . . drug paraphernalia, knowing, or under circumstances where one reasonably should know, that it will be used to . . . inject . . . or otherwise introduce into the human body a controlled substance. Any person who violates this subsection is guilty of a misdemeanor." (citing relevant portions of WASH. REV. CODE § 69.50.412(2) (West 1995))).

[53] *Brockett*, 839 P.2d at 328. For the Omnibus AIDS Act in its entirety, see WASH. REV. CODE § 20.24 (West 1995). For the particular provisions upon which the court relied in approving the Spokane needle exchange, see WASH. REV. CODE §§ 70.24.400(3)(b)(v), 400(12) (West 1995).

[54] *Brockett*, 839 P.2d at 327 (noting that "[u]nder the AIDS act, the largest county in each region is directed to develop a service 'plan' which meets listed statutory requirements."). The plans are to include, inter alia, "intervention strategies to reduce the incidence of HIV infection among high-risk groups, possibly including needle sterilization and methadone maintenance." WASH. REV. CODE § 70.24.400(3)(b)(v) (West 1995). The statute also authorizes the plans to make use of "the appropriate materials . . . in the prevention of control of HIV infection." WASH. REV. CODE § 70.24.400(12) (West 1995).

[55] WASH. REV. CODE §§ 70.05.060, .070 (West 1995).

constitutional right . . . ."[56] The court emphasized the extraordinary power delegated by the state legislature to cities and health boards in their efforts to control contagious diseases and concluded that the needle programs adopted under the Omnibus Act were valid notwithstanding the conflicting paraphernalia distribution statute.[57] The court also noted that such broad statutory delegation of police powers in general, and of public health powers in particular, is authorized under the Washington Constitution.[58] Finally, the court gave significant weight to the approval of local needle exchanges by both the State Board of Health and the State Department of Health.[59]

C. *New Jersey*

The operators of an unapproved needle exchange in New Jersey were less successful in defending their actions against criminal charges. The defendants in *State v. Sorge*[60] were charged with violating a New Jersey statute, under which it is "unlawful for a person to have under his control or possess with intent to use a hypodermic syringe, hypodermic needle, or any other instrument adapted for the use of a controlled dangerous substance . . . or to sell, furnish or give to any person such syringe, needle or instrument . . . ."[61] They moved to dismiss the case as a de minimis infraction under a state statute[62] which permits judges to dismiss prosecutions if the defendant's conduct "[d]id not actually cause or threaten the harm or evil sought to be prevented by the law defining the offense or did so only to an extent too trivial to warrant the condemnation of conviction,"[63] or if the conduct "presents such other extenuations that it cannot reasonably be regarded as envisaged by the Legislature in forbidding the offense . . . ."[64]

---

[56] *Brockett*, 839 P.2d at 329 (citation omitted).
[57] *Id.*
[58] *Id.* at 328 (citing applicable provisions of Washington Constitution: "Any county, city, town or township may make and enforce within its limits all . . . local police, sanitary and other regulations as are not in conflict with general laws." WASH. CONST. art. XI, § 11).
[59] *Brockett*, 839 P.2d at 332.
[60] 591 A.2d 1382 (N.J. 1991).
[61] N.J. STAT. ANN. § 2C:36-6 (West 1994).
[62] *Sorge*, 591 A.2d at 1383–84.
[63] N.J. STAT. ANN. § 22C:2-11(b) (West 1994).
[64] *Id.* § 22C:2-11(c).

The court denied the motion to dismiss, asserting that the statutory provision regarding de minimis infractions applies only to conduct so trivial that application of the Criminal Code would be absurd.[65] The court held that the defendants' actions facilitated illegal drug use and exacerbated the costs of illegal drug trade in contravention of the state's zero tolerance drug policy.[66] Recognizing arguments both for and against needle exchange as public policy,[67] the court deferred to New Jersey's "clear policy . . . to discourage illegal drug use."[68] The court noted that the balancing of interests involved in determining whether to permit needle exchanges "is quintessentially a legislative function" outside the discretion granted to courts under the de minimis infractions statute.[69]

D. *New York*

In New York, governmental reaction to needle exchange fluctuated substantially during the late 1980s and early 1990s. New York's Penal Law prohibits "knowingly and unlawfully possess[ing] or sell[ing] a hypodermic syringe or hypodermic needle";[70] however, the New York State Commissioner of Public Health has the discretion to authorize, by waiver, access to paraphernalia for the protection of the public health.[71] Between 1988 and 1990, a legally sanctioned syringe exchange operated in New York City under a public health waiver.[72] That program was terminated by former Mayor Dinkins shortly after his election.[73]

In 1991, a New York County judge acquitted defendants charged with criminal possession of hypodermic needles which were intended for needle exchange-style distribution.[74] Although the defendants were engaged in needle exchange activities without benefit of waiver and therefore in contravention of the Penal

---

[65] *Sorge*, 591 A.2d at 1384–85.
[66] *Id.* at 1385.
[67] *Id.* at 1385–86.
[68] *Id.* at 1386.
[69] *Id.* at 1386–87.
[70] N.Y. PENAL LAW § 220.45 (McKinney 1995).
[71] N.Y. PUB. HEALTH LAW § 3381(4) (McKinney 1995).
[72] Denise Paone et al., *Operational Issues in Syringe Exchanges: The New York City Tagging Alternative Study*, 20 J. COMMUNITY HEALTH 111, 113 (1995).
[73] *Id.*
[74] People v. Bordowitz, 588 N.Y.S.2d 507 (N.Y. City Crim. Ct. 1991).

Law, they successfully evoked the necessity defense.[75] New York's necessity defense statute permits emergency measures

> to avoid an imminent public or private injury which is about to occur by reason of a situation occasioned or developed through no fault of the actor, and which is of such gravity that, according to ordinary standards of intelligence and morality, the desirability and urgency of avoiding such injury clearly outweigh the desirability of avoiding the injury sought to be prevented by the statute defining the offense in issue.[76]

In regard to needle exchange, the judge in *Bordowitz* held it "reasonable for the defendants to believe their action necessary as an emergency measure to avert an imminent public injury."[77]

In the wake of the *Bordowitz* decision and mounting public pressures, the New York State Department of Health adopted regulations that expressly authorized needle exchange programs.[78] These regulations permitted the possession and furnishing of paraphernalia "in connection with the distribution or collection of hypodermic . . . needles for the purpose of preventing the transmission of human immunodeficiency virus in users of injectable drugs."[79] In 1992, a state court held that the regulations cannot be evoked as a defense against paraphernalia possession if the circumstances negate any inference that the defendant planned to exchange the needles for new, marked needles under a needle exchange program.[80] Nonetheless, needle exchange stands on a stable statutory foundation in New York.

E. *Massachusetts*

Massachusetts has rejected the necessity defense adopted in New York in *Bordowitz*.[81] In *Commonwealth v. Leno*,[82] the Supreme Judicial Court of Massachusetts upheld the conviction of persons operating a needle exchange. The defendants in *Leno* were prosecuted for operating a program[83] in violation of a Massachusetts statute requiring a prescription for the distribution of

---

[75] *Id.*
[76] N.Y. PENAL LAW § 35.05(2) (McKinney 1995).
[77] *Bordowitz*, 588 N.Y.S.2d at 511.
[78] N.Y. COMP. CODES R. & REGS. tit. 10, § 80.135 (1993).
[79] *Id.*
[80] People v. Monroe, 593 N.Y.S.2d 742 (1992).
[81] 588 N.Y.S.2d 507 (N.Y. City Crim. Ct. 1991)
[82] 616 N.E.2d 453 (Mass. 1993).
[83] Specific charges were "(1) unauthorized possession of instruments to administer

hypodermic needles.[84] They purchased new needles legally in Vermont, brought them into Massachusetts, and operated a free, grassroots needle exchange on a specific street corner during regular hours.[85]

At trial, the defendants admitted violating the statute prohibiting distribution of needles without prescription, but claimed justification by virtue of the defense of necessity.[86] The trial court judge refused to instruct the jury on the necessity defense, and on appeal the defendants claimed this refusal constituted error.[87]

The Massachusetts Supreme Judicial Court upheld the trial judge's determination, observing that the defense of necessity can be evoked only when "the danger motivating . . . unlawful conduct is imminent," and when the danger is clear rather than speculative.[88] Necessity, the Court observed, is limited to the purpose of exonerating those who commit crimes under high-pressure circumstances, in which the imminence of the danger renders an immediate assessment of competing harms extremely difficult.[89] The defendants' acts were held not to meet these criteria, but rather to be an attempted prevention of a remote future harm that was debatable and speculative rather than clear and imminent.[90]

F. *Federal Law*

Because direct public health policy traditionally falls within the states' police powers, federal laws pertinent to needle exchange have focused primarily on the more indirect issue of funding—specifically, on whether local needle programs can be supported with federal dollars. In 1988, Congress amended the Public Health and Welfare Act, creating a ban on the federal funding of needle exchange activities.[91] The 1988 Amendment,

---

controlled substances, and (2) unlawful distribution of an instrument to administer controlled substances . . . ." *Id.* at 454

[84] MASS. GEN. L. ch. 94C, § 27 (1990).

[85] *Leno*, 616 N.E.2d at 454–55.

[86] *Id.* at 455.

[87] *Id.* at 456.

[88] *Id.*

[89] *Id.*

[90] *Id.*

[91] Public Health and Welfare Act of 1988, Pub. L. No. 100-607, 102 Stat. 3048 (1988) [hereinafter the 1988 Amendment] (codified as amended in scattered sections of 42

however, provides an escape clause that can be evoked to permit needle exchange funding. Under this clause, federal funds can be used to finance needle exchanges if the Surgeon General determines that a program reduces both drug abuse and HIV transmissions.[92]

In 1993, researchers at the University of California, San Francisco suggested that the time had come to end the ban on federal funding.[93] They released a lengthy report (UCSF study), compiled for the CDC, from which the CDC concluded "the ban on federal funding of NEPs should be lifted to allow communities and states to use federal funds to support NEPs [Needle Exchange Programs] as components of comprehensive HIV prevention programs."[94] The U.S. Department of Health and Human Services (HHS) commissioned the CDC to review the study, but has not released the reviews publicly.[95] Some government health officials have evinced skepticism regarding the effectiveness of needle exchanges in the reduction of HIV, based in part on purported data from unpublished studies.[96] Others have suggested that the UCSF study does not make the case for the social utility of needle exchange with the degree of rigor necessary for a change in the law.[97] Peter Lurie, principal author of the UCSF study, responded with accusations that the present Administration is "cowering behind a legalistic fig leaf while thousands of drug users, their sex partners and their children are getting infected each year."[98]

U.S.C.). The ban on funding was reconsidered in the early 1990s. Stephen Burd, *U.S. Officials Reconsider Ban on Research into Needle-Exchange Programs*, CHRON. HIGHER EDUC., Dec. 18, 1991, at A27. As of this writing, the ban on federal funding of needle exchanges remains intact.

[92] The text of the statute states:

> None of the funds provided under this Act or an amendment made by this Act shall be used to provide individuals with hypodermic needles or syringes so that such individuals may use illegal drugs, unless the Surgeon General of the Public Health Service determines that a demonstration needle exchange program would be effective in reducing drug abuse and the risk that the public will become infected with the etiologic agent for acquired immune deficiency syndrome.

42 U.S.C. § 300ee-5 (1994).

[93] *Government Researcher Says U.S. and CDC Covering Up Needle-Exchange Studies*, AIDS WKLY., Mar. 6, 1995.

[94] *Id.*

[95] *Id.*

[96] *Id.*

[97] *Released Report Says Needle Exchanges Work*, 273 JAMA 980, 981 (1995).

[98] *Id.*

While this statement may be true, the findings of the University of California study have not to date been considered sufficient to meet the high standard established for a Surgeon General override of the funding ban established by the 1988 Amendment.[99] As long as that Amendment stands, the Surgeon General's options are limited. Even if the Surgeon General ignores any HHS skepticism of the University of California study and sides with the CDC position, an override of the funding prohibition still requires the unlikely finding that needle programs reduce drug abuse. Accordingly, needle exchange programs are likely to receive federal funding only if Congress revokes the 1988 Amendment.

II. Policy Perspectives Regarding Needle Exchange Programs

During the past several years, many reasoned and impassioned assertions for and against needle exchange have appeared in the press. The arguments on both sides are frequently supported by insufficient evidence. The public is left with plausible rhetoric but a dearth of proof. Nonetheless, it is important to begin any analysis of needle exchange by understanding the positions of both supporters and critics. These perspectives serve two important functions: (1) they shape the policy debate, and (2) they help formulate testable hypotheses that are relevant to the debate.

This section will examine the conceptual arguments most frequently utilized by both proponents and opponents of needle exchange programs.

---

[99] From a practical standpoint, the escape clause provides a threshold that is nearly impossible to meet and imposes an unduly heavy burden. A Surgeon General override requires not only a finding that needle exchanges are effective at curbing the spread of AIDS, but also concrete evidence that the programs reduce drug abuse. This latter requirement is extraneous to the primary goal of needle exchange programs: reduction of the spread of AIDS through needle sharing. A program that curbs HIV transmissions is a success in itself. It is unreasonable to expect that it will effect a decrease in drug abuse as well. Because the standard is overly burdensome, the Surgeon General override provision is unlikely ever to be invoked.

A. *The Arguments for Needle Exchange*

Proponents contend needle exchange programs can reduce the incidence of future HIV transmission[100] by giving injection drug users access to clean needles[101] and encouraging the disposal of used needles.[102] If injection drug use remains constant,[103] a shift away from needle sharing should function prophylactically and result in a decline in the spread of AIDS attributable to drug injection. Needle exchange supporters assert that the preventive value of needle exchange programs[104] outweighs any increase in illegal drug use. While some advocates suggest that needle programs do not increase drug use rates,[105] others go further and argue needle programs can decrease drug abuse rates by referring addicts to sources of treatment.[106]

---

[100] *See* Barry Brown, *AIDS Taking Toll on Toronto Males: Found to be No. 1 Cause of Death of Men Ages 25 to 54*, BUFFALO NEWS, Mar. 12, 1995, News, at 5 (noting observation of Parry Kendal, a recently retired medical officer, that the publicly funded Toronto needle exchange has been responsible for holding the HIV infection rate among injection drug users in the city to six percent).

[101] As needle sharing behaviors diminish, seroconversion rates decline. *See* Don C. Des Jarlais et al., *AIDS Risk Reduction and Reduced HIV Seroconversion Among Injection Drug Users in Bangkok*, 84 AM. J. PUB. HEALTH 452 (1994).

Of course, access to clean needles would have little value in the absence of a desire among intravenous drug users to protect their health. While some contend that intravenous drug users don't care about preserving their health and regulating their behavior accordingly, a recent study of drug outreach efforts contests this position. Robert S. Broadhead & Douglas D. Heckathorn, *AIDS Prevention Outreach Among Injection Drug Users: Agency Problems and New Approaches*, 41 SOC. PROBS. 473, 476 (1994).

[102] Commentators have supported needle programs on the basis of a variety of disciplinary perspectives. One prominent legal scholar has suggested that economic analysis favors distribution of clean needles. William N. Eskridge, Jr., *Review Essay: A Social Constructionist Critique of Posner's Sex and Reason: Steps Toward a Gaylegal Agenda*, 102 YALE L.J. 333, 339 (1992) ("Since AIDS is often, and increasingly, spread by intravenous drug users' sharing of contaminated needles and then, by heterosexual intercourse, to the drug users' sexual partners, economic analysis would . . . encourage strategies such as the distribution of latex condoms to the general population and of clean needles to those addicted to intravenous drugs."). Others have suggested an environmental nexus, contending that needle exchanges reduce the risks of improper needle disposal and transmission through accidental contact, as when a beach-goer steps on a contaminated needle. Kamrin T. MacKnight, *The Problems of Medical and Infectious Waste*, 23 ENVTL. L. 785, 807 (1993).

[103] This premise assumes, as do needle exchange proponents generally, that needle exchange does not cause injection drug use to increase.

[104] The ostensible prophylactic value of needle exchange is sometimes posited using anecdotal information regarding purported salutary effects. *See, e.g., Injecting Sense into Drug Treatment*, ECONOMIST, Aug. 29, 1992, at 56 (explaining that in the wake of needle exchange implementation in Edinburgh, "[d]iscarded syringes no longer gather in the wind-swept corners of the city's bleak council estates."). Needle exchange as prophylaxis has also been examined in numerous studies. For a sample of these, see *infra* part III.

[105] *See infra* note 355–362 and accompanying text.

[106] Joanne Jacobs, *The AIDS Breakout*, BALTIMORE SUN, Mar. 21, 1995, at 11A.

Needle exchange proponents often base their arguments on a realist position, suggesting that no alternatives have a comparable potential to save lives. These advocates reject counter-proposals like a stepped-up war against drugs and heightened educational or media campaigns as unrealistic or inadequate substitutes for needle programs.

Needle exchange proponents suggest that stepping up the war on drugs is an unrealistic approach to preventing HIV transmission. These advocates contend that anti-drug policies are unlikely to reduce illegal drug use.[107] If a war on drugs does not significantly reduce drug use within the next ten or twenty years, it will not impede the spread of AIDS. Accordingly, needle exchange supporters urge legislators to recognize substantial drug abuse as an unavoidable reality for the near-future time horizon critical to AIDS prevention policies.[108]

Because the law-and-order philosophy of the war on drugs has ostensibly been ineffective, some scholars posit "harm reduction" approaches[109] that seek to reduce the negative impact of drug abuse upon both users and society as more workable alternatives.[110] Some harm reduction proponents suggest that needle exchanges, which encourage sanitary practices among those who will inevitably continue to inject, are the only way to fill the AIDS-prevention gap left by an ineffective war against drugs.[111]

Needle exchange proponents view the position advocating an exclusive reliance on educational and media campaigns[112] as an

---

[107] *See* Roger E. Olson, *Need for Needle Exchange*, STAR TRIB., Apr. 6, 1995, at 16A ("[N]o war on drugs is going to eliminate intravenous drug use tomorrow or even next year. This means there will be people injecting drugs and if they don't have access to clean needles, they will be more likely to pick up the HIV virus.").

[108] As Osborn has observed, drug treatment facilities are not sufficiently abundant in the United States to meet the needs of drug addicts. She cites New York City as a "worst case example," in which fewer than 25,000 treatment slots are available for an estimated 200,000 heroin addicts. June E. Osborn, *AIDS: Politics and Science*, 318 NEW ENG. J. MED. 444, 446–47 (1988). Given such disparities between the challenge and the available resources, the war against drugs cannot realistically serve the immediate needs of AIDS prevention.

[109] For further discussion of harm-reduction approaches, and comparison to alternate approaches, see *infra* notes 146–148, 398–403 and accompanying text.

[110] *See* Don C. Des Jarlais & Samuel R. Friedman, *AIDS and the Use of Injected Drugs*, SCI. AM., Feb. 1994, at 82, 88 ("The harm-reduction perspective pragmatically acknowledges the difficulties of ending all misuse of psychoactive drugs. Nevertheless, it emphasizes the possibilities for reducing the individual and social harm associated with drug use.").

[111] *Id.*

[112] Some educational programs have been as controversial as needle exchange programs. For example, a videotape funded by the public health department of Toronto at a cost of approximately $2,000 was criticized for ostensibly providing "tips for addicts"

ineffective alternative to needle exchange. While advocates of an educational and media campaign claim that high-quality education can eliminate AIDS if it is accurate and thoroughly pervasive, needle exchange proponents point out that this position incorrectly assumes that recipients of information regarding the ways in which HIV is transmitted will act rationally and adopt abstinence as a strategy against contamination. The argument in favor of education presumes people will alter their behaviors and avoid high-risk activities, provided they know (1) the nature of those activities, and (2) the optimal means of modifying their behavior. A ubiquitous understanding of means of transmission and optimal avoidance strategies logically combines with perfect rationality to eliminate transmission associated with high-risk behaviors. Under this model, needle exchange programs would inefficiently replicate the function of perfectly effective educational programs.

Needle exchange proponents suggest that the reasoning of this theory is fundamentally flawed; the inadequacy of educational and media campaigns is a function of two basic problems: (1) No campaign can transcend all logistical and cognitive impediments to perfect, ubiquitous dissemination; and (2) High demand for a constrained supply of needles increases the likelihood that even informed drug users will continue to share needles. These problems are addressed separately below.

1. Logistical and Cognitive Impediments to Perfect, Ubiquitous Dissemination

The notion of an educational program pervasive enough to supplant the need for needle exchange is unrealistic.[113] Even if everyone who received the relevant information could and would alter their risky behaviors, many injection drug users would never receive the necessary information. At least two sources of

---

without sending any messages encouraging people to stop using drugs. The tape has been sold to needle exchanges and public health agencies for $35. Leslie Papp, *Taxes Let Addicts Make Needle Video; Step-by-Step Instructions "Minimize" Harm*, TORONTO STAR, Mar. 15, 1995, at A7.

[113] This is not to suggest that educational and media campaigns are ineffective, but rather that they cannot realistically be expected to be one hundred percent effective at influencing behavior. As one commentator observes, "mass media campaigns directed at behavioral change for prevention of HIV infection must continue." Ilise L. Feitshans, *Foreshadowing Future Changes: Implications of the AIDS Pandemic for International Law and Policy of Public Health*, 15 MICH. J. INT'L L. 807, 821–22 (1995) (book review).

imperfect educational dissemination exist—logistical sources and cognitive sources. To date, there is no reliable mechanism to ensure that any information will be disseminated throughout an entire population. This logistical difficulty is exacerbated because the population concerned is marginalized,[114] and therefore relatively poorly connected to the media sources upon which information campaigns typically rely.[115] Cognitive sources of imperfect dissemination are associated with the mind-altering effects of some drugs. As various forms of illegal drugs undermine the rational capacity of drug users, some who are exposed to information will not process it effectively. Needle programs have the potential to serve as a safety net for those who don't receive instruction and those who receive but fail to internalize instruction.

2. The Effects of Supply and Demand on Needle Sharing Behavior

High demand for needles and the constrained supply of new needles increase the likelihood that even informed drug users will continue to share needles. Because knowledge is not tantamount to rationality,[116] education of the public regarding the activities that increase the risk of contracting AIDS is, at best, an imperfect hedge against future infections. The link between knowledge of risks and modification of behavior is tenuous and fragile.[117] Drug users continue to share needles despite knowl-

---

[114] Along these lines, researchers refer to "hard-to-reach" audiences for AIDS awareness information, which may include young adults, members of minorities, the underserved, and the non-English speaking. Robert J. Donovan et al., *Paid Advertising for AIDS Prevention—Would the Ends Justify the Means?*, 106 PUB. HEALTH REP. 645 (1991). For discussion of one marginalized group in the context of injection drug use and AIDS, see Herman Joseph & Hilda Roman-Nay, *The Homeless Intravenous Drug Abuser and the AIDS Epidemic*, *in* AIDS AND INTRAVENOUS DRUG USE: FUTURE DIRECTIONS FOR COMMUNITY-BASED PREVENTION RESEARCH, National Institute on Drug Abuse Monograph 93 (C.G. Leukefeld et al. eds., 1990).

[115] For instance, injection drug users living on the streets are unlikely to be exposed to televised ad campaigns regarding the means of spreading AIDS.

[116] For the classic support of this assertion, see JAMES G. MARCH & HERBERT A. SIMON, ORGANIZATIONS 137–72 (1958).

[117] Dublin needle exchange founder Joe Barry argues that media campaigns must be supplemented with ground intervention in order to be effective: "You can have education messages and you can have mass media campaigns but you do actually need intervention at an individual level . . . . When did you ever change something that you do from looking at a leaflet? Okay, you get information from a leaflet but you change from internalising it, or hearing it from somebody else." Nuala Haughey, *"It Couldn't Happen to Me": A Plague Like AIDS Epidemic Hasn't Happened Yet Among the*

edge of the risks.[118] If educated drug users continue to engage in behaviors likely to spread AIDS, then education alone cannot be an adequate control.[119]

Initially, the irrationality of risky behavior among educated persons may appear puzzling. After all, the stakes of needle sharing, exposure to a highly debilitating disease[120] and premature death,[121] appear overwhelming relative to the fleeting euphoria associated with drug use. Yet, poverty and addiction explain needle sharing among educated drug users.[122] Illegal drug use is disproportionately a problem among the poor,[123] who are financially ill-equipped to purchase new needles. In addition, many forms of drug use are addictive.[124] Under the influence of sometimes extreme physiological and psychological craving,[125] it is

---

Heterosexual Community but HIV Infection is a Growing Problem Here, IRISH TIMES, Mar. 20, 1995, Well and Good, at 10.

[118] For discussion of some of the characteristics of substance abuse populations that might explain this phenomenon, see Sari H. Dworkin & Lester Pincu, *Counseling in the Era of AIDS*, 71 J. COUNS. & DEVEL. 275, 279 (1993); Reese M. House & Catherine M. Walker, *Preventing AIDS via Education*, 71 J. COUNS. & DEVEL. 282, 283 (1993).

[119] For discussion of some of the difficulties associated with behavioral changes among injection drug users, see Don C. Des Jarlais & Samuel R. Friedman, *The Psychology of Preventing AIDS Among Intravenous Drug Users: A Social Learning Conceptualization*, 43 AM. PSYCH. 865 (1988).

[120] *See, e.g.*, James D. Adams et al., *Pneumocystis Carinii Pneumonia in HIV Infected Patients: Effects of the Diseases on Glutathione and Glutathione Disulfide*, 24 J. MED. 337, 338 (1993) (describing pneumocystis carinii pneumonia as one such "debilitating disease of the lung that can accompany HIV infection").

[121] Harold W. Jaffe, *AIDS: Epidemiological Features*, 22 J. AM. ACAD. DERMATOL. 1167, 1168 (1990).

[122] A reduction in HIV transmissions in the gay male community has been attributed in part to educational and media efforts. *See CDC Aquired Immunodeficiency Syndrome (AIDS) Monthly Surveillance Report, supra* note 17. Educational programs aimed at gay men may have a better chance of being effective than those aimed at needle sharers, for whom addiction and poverty may be complicating factors.

Note, however, that transmissions have declined rather than ceased in the gay male community, even among the wealthy and the addiction-free. Ultimately, this means that any presumed human rationality will not be sufficient to curb what may appear to be short-sighted exchanges. Some persons will continue to risk their lives for fleeting sexual or drug-related pleasure even after learning that the price may be their health or life. This observation is simply a specification of the obvious generalization that human activity is often short-sighted, irrational, or driven by immediate urges rather than sober consideration of long-term costs and benefits. However pervasive AIDS prevention education may be, a number of persons will continue to become infected by exposure to HIV via high-risk activities.

[123] *See* Maureen M. Black et al., *Parenting and Early Development Among Children of Drug-Abusing Women: Effects of Home Intervention*, 94 PEDIATRICS 440, 440 (1994) (arguing that drug abuse often occurs in the midst of poverty).

[124] *See* Terry E. Robinson & Kent C. Berridge, *The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction*, 18 BRAIN RES. REVS. 247 (1993).

[125] *Id.*

not surprising that drug users share needles, despite an awareness of the associated risks.

The problem is exacerbated as the high demand for needles associated with addiction is compounded by supply constraints imposed as part of the war on drugs. Regulations that restrict access to syringes and needles[126] increase the likelihood that desperate users will share, circulate, and reuse needles.[127] Black market prices rise, further inhibiting access to new paraphernalia.[128] "Shooting galleries," typically abandoned buildings where drug users congregate to avoid arrest, further facilitate needle sharing activity.[129]

Needle shortages have led to deceptive black market practices, the harmful effects of which cannot be rectified by educating consumers. Street vendors sell used and repackaged needles as new products to desperate users,[130] undermining even informed purchasers' desire to protect their health.

All of these dynamics tend to create an environment in which educated drug users face access-related constraints in their efforts to adopt safer injection practices.[131] Accordingly, needle exchange supporters consider even the best-intended substitutes to be inadequate responses to the challenge of reducing HIV transmission rates.

B. *The Arguments Against Needle Exchange*

The idea that educational and anti-drug programs are inadequate to curb the spread of AIDS through needle sharing is intuitively reasonable. Yet, while the limitations of these programs are apparent,[132] the value of needle exchange as an appro-

---

[126] For discussion of these regulations, see *supra* notes 26–28 and accompanying text.

[127] Studies reveal that between 46% and 66% of needle sharers inject with used needles because clean needles are unavailable or unobtainable at the time of injection. *See* P.A. Selwyn et al., *Knowledge about AIDS and High-Risk Behavior among Intravenous Drug Users in New York City*, 1 AIDS 247 (1987); *See also* S. Paine et al., *Letter, AIDS in Drug Abusers*, 143 MED. J. AUSTL. 631 (1985).

[128] Andrew D. Firlik & Katrina Schreiber, *AIDS Prevention by Needle Exchange*, 92 N.Y. ST. J. MED. 426, 426–27 (1992).

[129] *Id.* at 426.

[130] *Id.* at 426–27.

[131] *See* Gerry V. Stimson, *Syringe-Exchange Programmes for Injecting Drug Users*, 3 AIDS 253, 253 (1989) ("Injectors commonly state that they share syringes because of difficulties in obtaining them.").

[132] Acknowledgement of these limitations is not meant to suggest that behavioral responses are unsusceptible to counseling interventions. Indeed, behavioral research is

priate supplement has been controversial.[133] Governmentally supported or approved needle exchanges have generally been adopted only after some public or political protest, particularly in the United States, where many of the approximately seventy exchange programs operate without legal sanction.[134] Other proposed needle exchanges have been delayed or thwarted by the objections of critics.[135]

Common-sense arguments in favor of needle programs are easily countered by common-sense arguments against them. Some critics, particularly opponents of federal funding of needle programs, contend that investment in needle exchange is a suboptimal use of scarce resources.[136] Others focus on the illegality of injecting proscribed drugs. For example, Christine Whitman, the governor of New Jersey, has stated that she is opposed to needle exchanges "because she 'doesn't want to be a party to an illegal activity.'"[137] Other opponents have responded with cynicism to the hypothesis that needle programs will reduce instances of needle sharing, thereby reducing net AIDS transmission through use of contaminated needles.[138]

---

capable of addressing ways in which interventions can alter high-risk behaviors more effectively. Jeffrey A. Kelly et al., *Psychological Interventions to Prevent HIV Infection are Urgently Needed: New Priorities for Behavioral Research in the Second Decade of AIDS*, 48 AM. PSYCH. 1023 (1993). Nonetheless, even the most optimally effective interventions cannot realistically be expected to evoke perfectly rational responsiveness among injection drug users.

[133] Opposition to needle exchange programs has come from a diverse array of sources. According to one report, "conservative leaders and those who have suffered most from the legacy of America's racial history have become unwitting allies" in the opposition to needle exchange. David L. Kirp & Ronald Bayer, *Needles and Race*, ATLANTIC, July 1993, at 38, 42.

[134] Charlie Rose Transcript # 1354-3, WNET Educational Broadcasting Company, Apr. 12, 1995 (citing statement of documentary filmmaker Vanessa Vadim). *See also Residents Use Own Money for Illegal Needle Exchange*, ORLANDO SENTINEL, Mar. 11, 1995, at A16 (reporting plans to implement an underground needle exchange in Atlanta).

[135] *See Needle Exchange Tabled in Contra Costa County*, S.F. CHRON., Apr. 26, 1995, at A14 (citing delay of proposed needle exchange by Contra Costa County supervisors, who were "swamped with calls and letters opposing the idea"); *see also* Jamie Schield, *Should County Distribute Condoms, Bleach Kits?*, DALLAS MORNING NEWS, Apr. 7, 1995, at 29A (noting and criticizing Dallas County Commissioners' recent decision to end all distribution of condoms and needle sterilization kits in county health programs).

[136] *See* Wayne Hearn, *Track Record*, AM. MED. NEWS, Apr. 17, 1995, at 8, 12 ("It's far too early for the government to be putting up large amounts of money for needle exchange that could be spent on better prevention and treatment strategies for drug addicts." (quoting Eric Voth, chair of the International Drug Strategy Institute)).

[137] *Clean Needles are Vital in the Fight Against AIDS*, RECORD (Bergen County, N.J.), Mar. 22, 1995, at C6.

[138] Linda W. Lockwood, *Public Needle Exchange Devastates Community*, L.A. TIMES, Mar. 5, 1995, at B18.

Needle exchange opponents suggest that the purported benefits are illusory and "just don't work as we are led to believe."[139] This could be true for several reasons. It is possible that drug users may fail to utilize needle exchange programs consistently, so that transmission via contaminated needles ultimately may not be reduced. Some critics also suggest that needle programs sanction and encourage the illegal use of drugs.[140] If this hypothesis is true, needle programs could be dysfunctional at two levels: they could escalate the drug abuse problem, and they could even increase the rate of HIV transmission.[141]

State public health laws requiring a prescription to obtain needles arguably set a tone, signaling that injection drug use is condoned only for authorized medical purposes. Some claim that diminishing this cultural tone through needle exchange programs would exacerbate numerous social ills. They contend that predicted increases in drug use are likely to harm minorities in disproportionate numbers, and one commentator went so far as to suggest that a proposed needle exchange program in New York City comprised a "genocidal campaign against black and Hispanic people."[142] A needle program widely considered to be successful—the pioneering Tacoma, Washington program—has come under attack for allegedly promoting crime.[143] On the local level, some have argued that needle exchange programs should

---

[139] *Id.*

[140] Opponents have included President Bush and his Administration's Department of Health and Human Services, which contended that needle exchange programs "can be viewed as sanctioning drug abuse." *Federal AIDS Panel Favors Needle Exchange*, AMER. PHARMACY, Oct. 1991, at 17.

[141] The idea that needle exchange programs could increase rates of HIV transmission uses the following reasoning: If needle programs increase drug abuse, then the total number of persons using injection drugs illegally obviously increases. With this increase, there is a concomitant increase in the number of potential needle sharers. Because needle exchange program users will not all utilize the exchanges consistently, the increase in injectors may increase needle sharing activity more than the program's distribution of clean needles will decrease needle sharing activity. If this is the case, needle programs would actually increase the rate of HIV transmission through use of shared needles.

[142] Michel Marriott, *Needle Exchange Angers Many Minorities*, N.Y. TIMES, Nov. 7, 1988, at B1. For further discussion of the response of some African-American leaders to needle exchange programs, see Ronald Bayer, *Commentaries: Rethinking Aspects of Aids Policies*, 11 J. CONTEMP. HEALTH L. & POL'Y 457, 461–63 (1995).

[143] Barbara Clements & Elaine Porterfield, *Tacoma Official Wants to Ban Needle Swap*, MORNING NEWS TRIB., Apr. 4, 1995, at A1 ("People will commit crimes to buy drugs and use these free needles to stick in their arms." (quoting needle exchange opponent, Councilman Steve Kirby)).

be rejected because they are likely to draw drug addicts into the community[144] and reduce its quality of life.[145]

Underlying these and other arguments against needle programs is a larger, ideological tension within which controversies over needle programs tend to be subsumed—what Raistrick calls "the pervasive tension between harm reduction strategies and measures designed to reduce drug taking."[146] Under this rubric, critics associate needle programs with other policies or practices that purportedly indulge or coddle drug users, such as the Swiss "Needle Park" experiment, in which a Zurich park was allocated for drug use.[147] These policies arguably connote an acceptance of drug abuse which might encourage undesirable behaviors. The fallacy of harm reduction, the critics reason, is in the notion that working with drug users, and not against them, will yield a net reduction in social harm.[148]

Specifically, critics suggest that needle programs either directly or indirectly encourage drug abuse in two ways—by creating the impression of public approval of the injection of contraband drugs,[149] and by providing the tools necessary to inject such drugs safely.[150] These two arguments will be addressed

---

[144] *See* Peter Y. Hong, *L.A. Elections/4th District: Ferraro Has Strong Advantage*, L.A. TIMES, Apr. 5, 1995, at B1 (mentioning city council candidate Linda W. Lockwood's rationale for fighting a Hollywood needle exchange program).

[145] *See* Monte Williams, *Neighborhood Report: Villages East and West; A Rough New Neighbor in a Rough Old Neighborhood*, N.Y. TIMES, Mar. 19, 1995, § 3, at 6 (reporting a citizen's response to blood freshly splattered on the steps of her house: "I keep a supply of bleach and ammonia on hand to clean up the blood . . . . They also defecate, urinate, and vomit right here," and observing that critics of the program question whether "needle exchange, however well-intentioned and effective, is too costly in terms of quality of life."); *see also* Elizabeth Hess, *The Malling of Soho: Where Will the Art World Go?*, VILLAGE VOICE, Mar. 14, 1995, at 23 (discussing the formation of the Soho Community Coalition, a function of which has been to keep a needle exchange program from locating in the Soho section of Manhattan).

[146] Duncan Raistrick, *Report of Advisory Council on the Misuse of Drugs: AIDS and Drug Misuse Update*, 89 ADDICTION 1211, 1211 (1994).

[147] For discussion of the Needle Park experience, see Christian Huber, *Needle Park: What Can We Learn from the Zurich Experience?*, 89 ADDICTION 513 (1994).

[148] Raistrick describes the possible fallacy of harm reduction approaches in terms of what he calls "harmful harm reduction." Specifically, he questions whether it is reasonable to view HIV as more dangerous than drug abuse. *See* Raistrick, *supra* note 146, at 1211–12.

[149] *See Clean Needles are Vital in The Fight Against AIDS*, RECORD (Bergen County, N.J.), Mar. 22, 1995, at C6 (noting that "government officials who oppose needle-exchange programs typically say they don't want to sanction illicit drug use").

[150] *See* Bob Curley, *Political Correctness Backfires in Drug Policy*, DRUG POL'Y, ALCOHOL & DRUG ABUSE WK., Apr. 17, 1995, at 5 (citing arguments of needle exchange opponents that provision of clean needles to drug addicts encourages drug use). *See also* Michael H. Shapiro, *Symposium: Regulation as Language: Communicating Values by Altering the Contingencies of Choice*, 55 U. PITT. L. REV. 681, 786

throughout this Article, respectively, as the "public sanction/wrong message" theory, and the "provision of tools" theory.

The public sanction/wrong message theory views both legalistic and cultural disapproval of contraband drug use as strong inhibitors which discourage at least some margin of potential users from ultimately deciding to experiment with drugs.[151] Supporters of the theory reason that needle exchange sends an ambiguous signal that dilutes the effectiveness of drug control laws. Some will receive this signal as government condonation or countenance—if drug injection is seriously forbidden, why would the government be giving people implements for the express and admitted purpose of injecting drugs?[152]

The provision of tools theory suggests that needle programs make the injection of illegal drugs feasible and even convenient. Particularly in states where the dispensing of paraphernalia is controlled by prescription laws[153] or by regulations prohibiting knowing provision for illegal use,[154] needle programs might facilitate the injection of drugs by enhancing access to the necessary implements.

If needle exchange programs engender even a small marginal increase in drug abuse, future opponents are likely to recommend any less harmful alternatives in lieu of needle exchange. One intriguing answer could lie in technology; indeed, the concept of non-reusable injection technology has recently begun to receive attention as a potential substitute for needle exchange programs.[155] While the potential of non-reusable needles and syringes to obliterate needle sharing opportunities is a desirable area for further study, Des Jarlais has suggested that redesigned injection equipment is "unlikely to reduce the spread of HIV and

---

(1994) ("The opposition to needle exchange and condom distribution often focuses on the risk that the disapproved behavior will increase in incidence because of the learning effects of the public distribution programs—programs that openly acknowledge certain practices and . . . further them by providing tools that facilitate the practice by making it safer.").

[151] *See* Maki Becker, *Clean Needles: Solution for an Epidemic?*, L.A. TIMES, Sept. 9, 1994, at B2.

[152] *See* Arthur Gould, *Sweden's Syringe Exchange Debate: Moral Panic in a Rational Society*, 23 J. SOC. POL'Y 195 (1994) (observing suggestion by Swedish needle exchange opponents that the programs may encourage rather than simply condone drug abuse).

[153] *See* CAL. BUS. & PROF. CODE § 4143 (West 1995).

[154] *See* CAL. HEALTH & SAFETY CODE § 11364.7 (West 1995).

[155] *See generally* PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON SELF-DESTRUCTING (NON-REUSABLE) SYRINGES: STRATEGIES FOR BLOCKING TRANSMISSION OF HIV, HEPATITIS AND OTHER BLOOD-BORNE PATHOGENS (M. Marmour ed., 1991).

may have other unintended consequences."[156] He observes that no truly non-reusable syringe has been designed to date, and that a black market for easily reusable paraphernalia would likely undermine the effectiveness of relying on difficult-to-reuse needles as an AIDS prevention program.[157] Des Jarlais also notes, "evidence indicates that many of the proposed redesigns would interfere with usual drug-taking practices, making many drug users unlikely to accept them."[158] Accordingly, he suggests that difficult-to-reuse hardware may be effective in limited situations and amongst targeted groups who are unlikely to attempt to defeat the equipment.[159] While these limitations may presently weaken arguments that an effective needle exchange is redundant and incurs unnecessary costs, future technological improvements may bolster the viability of replacing the needle exchange concept with self-destructing needles.

The implementation of some needle programs may be delayed or halted in light of any of the aforementioned challenges, while other programs are reluctantly and provisionally accepted[160] under the proviso that they be conducted experimentally, as controlled clinical trials.[161] Even those programs that are designed for research purposes, or with data collection components aimed at answering difficult questions regarding the effects of needle programs, have encountered serious, influential opposition.[162] The controversy over needle programs invariably converges upon the same basic issues—the extent, if any, to which needle programs reduce the transmission of HIV; the nature and characteristics of needle programs that appear most likely to be effective; and the extent, if any, to which needle programs increase the incidence of illegal drug use.[163] These three areas of inquiry, critical to the

---

[156] U.S. CONGRESS, OFFICE OF TECHNOLOGY ASSESSMENT, DIFFICULT-TO-REUSE NEEDLES FOR THE PREVENTION OF HIV INFECTION AMONG INJECTING DRUG USERS—BACKGROUND PAPER OTA-BP-H-103 (1992).

[157] *Id.*

[158] *Id.*

[159] *Id.*

[160] Chris A. Raymond, *U.S. Cities Struggle to Implement Needle Exchanges Despite Apparent Success in European Cities*, 260 JAMA 2620 (1988).

[161] For detailed discussion of one such example in New York City, see Warwick Anderson, *The New York Needle Trial: The Politics of Public Health in the Age of AIDS*, 81 AM. J. PUB. HEALTH 1506 (1991).

[162] For discussion of both prominent supporters and detractors of experimental needle exchange programs, see Harold M. Ginzburg, *Needle Exchange Programs: A Medical or a Policy Dilemma?*, 79 AM. J. PUB. HEALTH 1350, 1351 (1989).

[163] These issues were addressed in a report issued by Britain's Advisory Council on the Misuse of Drugs, which concluded that "HIV is a greater threat to public and

evaluation of needle exchange proposals, are addressed respectively in Parts III through V.

III. The Literature on Needle Exchange Program Effectiveness in Reducing New Transmissions of HIV

Arguments based on raw emotion and unsubstantiated intuition tend to dominate the needle exchange debate. These conflicting arguments frequently embrace mutually inconsistent suppositions or conclusions and are rarely substantiated by hard evidence. In evaluating whether to support needle exchange, the onus is on lawmakers to exercise great care in sifting through the rhetoric and to provide an environment in which further experimentation can flourish. Arguments on each side of the controversy must be tested to determine which competing but incompatible claims have the most merit.

Needle exchange programs should be approached much like promising but untested drug treatments.[164] Programs should be developed provisionally and under experimental auspices so investigators can examine their impact on infection rates and their possible side effects.[165] Dependence on political rhetoric, hyperbole, and overwrought predictions of disaster can only lead to poor public policy.[166] Instead, studies are needed so lawmakers can understand how needle programs function and the variables that will affect their success or failure.

As shall be evident in the examination of the research on needle programs to date which follows, many questions remain unsettled due to the scarcity of information and conflicting findings. For example, while some studies have found needle exchange

---

individual health than drug misuse." Notwithstanding this finding, Britain's Department of Health subsequently declined to adopt a needle program. One source suggested that the decision was based on governmental fear of "being seen as condoning drug misuse." Steve Connor, *Advisors are Bitter at AIDS Ruling*, New Scientist, Apr. 7, 1988, at 17.

[164] For detailed discussion of the processes used to evaluate previously untested drugs for HIV and AIDS treatment, see Steven R. Salbu, *Regulation of Drug Treatments for HIV and AIDS: A Contractarian Model of Access*, 11 Yale J. on Reg. 401 (1994).

[165] As in the case of drug testing, side effects would refer to potential ramifications upon other parts of the system. As a drug effective in the treatment of a heart condition may have negative side effects through increased susceptibility to headaches, so a needle program that reduces transmission of HIV may have negative side effects through increased illegal drug use. Pilot programs could measure both effect upon transmission rates and other unintended social and economic ramifications.

[166] For discussion of how the political process ended the life of a pilot program in New York City, see Anderson, *supra* note 161.

attenders to be less likely to pass on needles than non-attenders,[167] other studies have observed exactly the opposite trend.[168] The absence of conclusive findings is exacerbated by the potential limitations of external validity or the generalizability of observations.[169] The findings of pilot needle studies like those discussed in this Part may be context-specific. Any number of variables across different populations may affect the utility of the programs and the characteristics of optimal programs.[170]

To date, research into the effectiveness of needle programs has provided limited insight, reflecting in part the dearth of data that results from the reluctance of many communities to adopt the programs. Scarcity of research findings may also be a function of logistical difficulties. The effect of needle programs on the transmission of HIV is difficult to measure. Comparisons of new infections believed to be associated with injection drug use, before and after implementation of needle programs, provide limited useful information. Such data reflect correlational trends from which causality generally cannot be inferred.[171] Because controls are notoriously difficult to employ in field settings,[172] changes in transmission rates are not easily attributable to the adoption of a needle program. Nonetheless, investigators have developed studies using various creative methodologies, in an effort to overcome these impediments. Some major findings, as well as the limitations of these findings, are discussed in the subsections that follow.[173] These are offered to help the reader

---

[167] *See, e.g.*, Martin Frischer & Lawrence Elliott, *Discriminating Needle Exchange Attenders from Non-Attenders*, 88 ADDICTION 681 (1993).

[168] H. Klee et al., *The Sharing of Injecting Equipment Among Drug Users Attending Prescribing Clinics and Those Using Needle-Exchanges*, 86 BRIT. J. ADDICTION 217 (1991).

[169] While external validity refers to a number of inferential research concerns, the aspect at issue here is the generalizability of findings across a wide variety of settings and conditions. For detailed discussion of external validity, see JOHN M. NEALE & ROBERT N. LIEBERT, SCIENCE AND BEHAVIOR: AN INTRODUCTION TO METHODS OF RESEARCH 164–91 (1973).

[170] The external validity problem is not peculiar to needle exchange research, but rather constrains virtually all scientific and quasi-scientific investigation. Our inability to make gross generalizations is a reflection of a diverse universe from which studies of small, particularized samples can represent phenomena that are highly common, highly context-specific, or some mix of these two extremes.

[171] Richard J. Hickey & I. Elaine Allen, *Surgeon General's Reports on Smoking and Cancer: Uses and Misuses of Statistics and of Science*, 98 PUB. HEALTH REP. 410 (1983).

[172] For discussion of difficulties in achieving control in field settings, see FRED N. KERLINGER, FOUNDATIONS OF BEHAVIORAL RESEARCH 370–75 (3d ed. 1986).

[173] While many investigators have studied needle exchange programs, some studies are more ambitious and helpful than others. Although the studies chosen for discussion

(1) to become familiar with our present understanding of the relationship between needle exchanges and HIV transmission, and (2) to develop an appreciation for the limits of our present knowledge, which is bounded by both methodological difficulties and imperfections in methodological execution. For purposes of dicussion, studies are classified by their methodology.

A. *Return Rate Studies*

One way to study the value of needle exchange is simply to quantify the effect such programs have on needle return rates. Some investigators have used return rates of needles distributed through needle exchange programs as a rough proxy for effectiveness in reducing HIV transmissions.[174] As Joseph Guydish et al. observed in their study of a San Fransisco needle program (the Guydish study), "[i]n the ideal program, 100 percent of needles distributed would be returned in a short time, implying consistent use of new needles and a low likelihood of needle sharing in the exchange population."[175]

Nonetheless, the data derived from needle return rate studies has not yielded dispositive results. The Guydish study found "nearly 50 percent of distributed needles returned within two weeks"[176] in a San Francisco study. The inference they derived from this finding was that "some stability [exists] in the exchange population, and that many patrons use the exchange repeatedly."[177] While this finding does show behaviors moving in the desired direction, it does not tell us whether transmission rates are reduced as a result. Even if studies were to uncover much higher rates of return, it is still possible (although admittedly unlikely) that HIV transmission rates could remain con-

---

in this part are intended to be representative of the better studies to date, the number of investigations necessitates that this sample not be exhaustive. In the process of selecting studies that have had an impact on the literature, some useful studies must inevitably be omitted.

[174] The use of this measure requires that an experimental program exchange new needles for used ones. While it is conceivable that a program might distribute clean needles without an exchange, this approach has not generally been adopted. Most needle programs are called "exchanges" because they require some form of used needle return. The degree to which programs enforce the requirement of one-for-one exchange varies among programs.

[175] Joseph Guydish et al., *Evaluating Needle Exchange: Do Distributed Needles Come Back?*, 81 AM. J. PUB. HEALTH 617 (1991).

[176] *Id.* at 618.

[177] *Id.*

stant or transmission delayed rather than reduced. For example, a needle's return within "a short time"[178] tells us only that the needle came back within the allotted time frame, i.e., within two weeks in the Guydish study.[179] The relatively speedy return of a needle does not, however, trace the history of the needle. The needle may have been used by multiple drug users at one sitting, or by any number of sequential drug users in a series of sittings.

Difficulty in tracking needle life, and therefore inferring program success from return rates, is exacerbated when logistical limitations of needle exchange programs undermine users' resolutions to avoid sharing. For example, the Guydish study exchanged needles only once a week during regular business hours.[180] While the program permitted exchange of up to ten needles each week,[181] the inflexibility and relative infrequency of exchange operations may have discouraged regular exchangers' exclusive reliance on unused needles.[182] In the end, substantial needle exchange returns within two weeks are promising behaviors, but are ultimately inconclusive measures of the effectiveness of the exchange concept.

Return rate studies are further limited by political, logistical, and methodological factors. In one D.C. study (the Vlahov study), Vlahov et al. were initially limited by the nature of the syringe exchange program that was approved by the City Council, which incorporated a number of concessions to the program's opposition.[183] The authors of the study observe that "program design was achieved only with compromise. The program was authorized to operate for only 60 days, with the impact to be reviewed by the council prior to further action."[184]

For two reasons, sixty days is simply too short a time to learn anything about the long-term impact of needle exchange. First, such brief time frames will generate data exclusively regarding

---

[178] *Id.*

[179] *Id.*

[180] *Id.* at 617.

[181] *Id.*

[182] These comments reflect the difficulties inherent in this type of research. The difficulty in tracking the use of a needle during its street-life would have posed challenges to researchers even if the investigators had decided to measure return rates after one week. The underlying problem would have remained largely unaltered. The potential for needle sharing in one week is not qualitatively distinguishable from the same potential over two weeks.

[183] David Vlahov et al., *A Pilot Syringe Exchange Program in Washington, D.C.*, 84 AM. J. PUB. HEALTH 303 (1994).

[184] *Id.*

short-term needle exchange impact, which is ultimately less important to law and policy makers than long-term impact. Second, an individual's consistent utilization of exchange programs to the exclusion of needle sharing practices may take time to develop due to a gradual learning process and a gradual development of trust between the individual and the exchange program staff.

In the face of political constraints, design of exchange program studies tends to suffer. The Vlahov study was able to enroll only thirty-three drug users in their study.[185] While the authors do not attribute the small sample specifically to political exigencies, they do observe that the City Council limited inclusion in the exchange program to "adult injection drug users residing in the district who had applied for entry into treatment for drug abuse and the city's central intake unit but were placed on a waiting list because treatment slots were full."[186] This restriction sets an artificial cap on the number of injection drug users available to be studied. It also restricts the generalizability of any results, because applicants to drug abuse treatment centers may have different motivational profiles and different behavioral responses to needle exchange than drug users in general. Ironically, the restriction resulted in the provision of information about the most promising group—those trying to stop using drugs—while disqualifying what is intuitively the most critical group—those not trying to stop using drugs, and therefore most at risk to continue engaging in unsafe practices.[187]

B. *Survey Research Studies*

Some studies have sought to understand behavioral changes among needle program participants by asking them questions about their behavior. This method of data collection falls under the heading of survey research. Among the most ambitious needle program survey studies to date is the analysis of Watters et al., which utilized eleven semi-annual surveys, from 1986 through 1992, in conjunction with an all-volunteer syringe exchange in

---

[185] Vlahov, *supra* note 183, at 303.

[186] *Id.*

[187] The authors of the study observe in their discussion of findings that critics "noted that restriction of eligibility to persons who already had made a decision to cease drug use seemed illogical, and that the persons most in need of clean needles were systematically excluded." *Id.* at 304.

San Francisco.[188] The program, known as "Prevention Point," was not legally sanctioned[189] but was tacitly supported by two successive mayoral administrations.[190] The program provided strict one-for-one exchange during regular evening hours and operated at street corner sites, as well as through mobile teams.[191] Unlike many needle program studies, the samples used by Watters et al. were impressive—the investigators had access to 5,644 useable surveys, ranging from approximately 400 to 600 surveys during each half-year reporting period.[192]

The Watters study documented impressive behavioral changes that suggest certain needle programs may yield meaningful reductions in transmission of AIDS among injection drug users. In the half year designated "Spring of 1992," Prevention Point exchanged 343,883 syringes, representing an average exchange of twenty-one syringes per client contact.[193] For the same spring 1992 period, forty-five percent of respondents said that they "usually" obtained syringes through the exchange.[194] The investigators also observed a dramatic decline in needle sharing behaviors over the life of the program: while 66.3% of respondents reported sharing needles in the spring of 1987, only 35.5% reported needle sharing by the spring of 1992.[195] Given independent findings that detected HIV-1 antibodies in seven percent of the syringes returned to Prevention Point,[196] some reduction in contamination opportunities can be reasonably attributed, ceteris paribus, to the Prevention Point program.[197] While the study is not conclusive evidence of a reduction in HIV transmission,

---

[188] John K. Watters et al., *Syringe and Needle Exchange as HIV/AIDS Prevention for Injection Drug Users*, 271 JAMA 115 (1994).

[189] For a discussion of legality of needle exchange programs under California law, see *supra* part I.A.

[190] Watters, *supra* note 188, at 116.

[191] *Id.*

[192] *Id.* at 115–16.

[193] *Id.* at 118.

[194] *Id.*

[195] *Id.*

[196] Joseph Guydish et al., *Detecting HIV Antibodies in Needle Exchange Syringes*, 6 AIDS 739 (1992).

[197] The ceteris paribus qualification is important here in light of the hypothesis that needle exchange programs increase illegal drug use. If that hypothesis is false and other variables remain unaffected by the needle exchange, Prevention Point certainly appears to diminish contamination opportunities. If the hypothesis is true or if other factors that have not been considered are negatively affected by the program, it is conceivable that observed diminished contamination opportunities are offset by other sources of increased contamination opportunity.

it does make a compelling case that the effect could be substantial.

In another survey study, van den Hoek et al. tracked 263 drug users in Amsterdam between December 1985 and April 1988, obtaining data from a combination of blood tests and interviews.[198] Volunteers for the study were recruited at six methadone outposts and one STD clinic, none of which were drug-free sites.[199] Blood samples and interview data were taken from participants at the start of the program and every four months thereafter, and participants were paid for each follow-up visit to encourage sustained participation in the study.[200] This incentive did not ensure regular follow-ups among all who started the program, and the investigators limited their analysis to persons who did return regularly.[201]

Van den Hoek et al. observed from interview responses that exclusive use of needle-exchange needles rose from 31.2% of participants at the first visit to 52.3% at the third visit, even though participant use of injection drugs remained stable.[202] Unfortunately, we cannot dismiss the possibility that those who remained in the study through the third visit were more responsible in general than those who had dropped out. Such a change in the composition of the sample attributable to attrition might account for some or all of the improvement observed over time.[203] The investigators also noted a sharp decline in the incidence of needle borrowing and lending over time.[204] Using a log linear model, however, they deduced that the reduction in borrowing needles and syringes was a study effect, unattributable to the needle program itself.[205]

The authors reported an increase in the use of the needle exchange system over time, which they attributed to the prevention campaign element of the needle program, rather than to any

---

[198] J.A.R. van den Hoek et al., *Risk Reduction Among Intravenous Drug Users in Amsterdam Under the Influence of AIDS*, 79 AM. J. PUB. HEALTH 1355 (1989).

[199] *Id.*

[200] *Id.*

[201] Regular returnees were selected from those returning within a period of between three and six months. Two hundred thirty-three participants returned for a second visit within the requisite time frame; 165 returned for a third visit within the requisite time-frame. *Id.*

[202] *Id.* at 1356.

[203] The sample for the first visit was 189, for the second visit, 156, and for the third visit, 107. *Id.*

[204] *Id.*

[205] *Id.*

intrusive effects of the study.[206] They also concluded that needle programs require intensive counseling components to be effective at maximizing total risk reduction.[207] Because the number of seroconversions tracked was too small to support a reliable comparative study of attack rates over time, the findings were based entirely on the survey responses and not on the more objective sequential HIV-test data.[208]

More recently, van den Hoek and other researchers updated their study, examining the behavior of injection drug users in Amsterdam from 1986 to 1992.[209] As in their earlier study, they found a consistent decline in borrowing and lending of injection equipment and in reuse of needles and syringes, and were forthright in their recognition of methodological and measurement limitations.[210] Like most findings in this area, their results are promising but inconclusive.

In the late 1980s, Hartgers et al. studied 145 volunteer subjects, all of whom were injection drug users in Amsterdam and who had injected during the preceding six months.[211] The volunteers were recruited from various sources,[212] so that behaviors of exchange users could be compared with behaviors of non-exchange users. The investigators used a two-stage interview process to accumulate data; of the 108 of the original 145 volunteers whom they tried to contact for follow-up, 60 participated in a second interview, an average of 13.5 months after the first.[213] The study gathered two types of information. First, demographic and behavioral history data were collected and entered as independent variables in a logistical regression model.[214] Second, various types

---

[206] *Id.* at 1357.

[207] *Id.*

[208] *Id.*

[209] Eric J.C. van Ameijdan et al., *Injecting Risk Behavior Among Drug Users in Amsterdam, 1986 to 1992, and its Relationship to AIDS Prevention Programs*, 84 AM. J. PUB. HEALTH 275 (1994).

[210] *Id.*

[211] Christina Hartgers et al., *The Impact of the Needle and Syringe-Exchange Programme in Amsterdam on Injecting Risk Behaviour*, 3 AIDS 571, 572 (1989).

[212] *Id.* at 572.

[213] Of the original 145 participants, only 108 were available to be contacted for a second interview because (a) 10 had originally been interviewed anonymously, (b) 2 had died, (c) 2 refused a second interview, and (d) 23 had returned to their country of origin. *Id.*

[214] The 13 independent variables were "age, sex, nationality, duration of injecting, frequency of injecting heroin the month before, frequency of injecting cocaine the month before, frequency of injecting the month before, amount of injecting now compared with 6 months before, frequency of injecting alone the month before, frequency of exchanging needles within the last six months, daily use of methadone,

of risk-related activities were observed as the dependent variables.[215]

The investigators compared the behaviors of needle exchange users with the behaviors of non-exchange users. They observed, "During the previous 6 months, 3% of exchangers and 27% of non-exchangers found themselves in the high-risk situation of possessing drugs but not clean needles daily."[216] Furthermore, "Eighty-two percent of the exchangers used their own needles only once, compared with 29% of the non-exchangers," and "[m]ore non-exchangers than exchangers had borrowed needles in the previous month as well as in the last 2 years."[217] From these and other behavioral observations, statistically controlled for alternative or competing explanations, the authors inferred that "the exchange programme seems to be effective in lowering the risk level of injecting . . . ."[218]

The work of Donoghoe et al. examines self-reported changes among participants in fifteen government-sponsored pilot syringe exchange programs in England and Scotland.[219] The study purports to measure "self-reported attitudinal, knowledge and behavioural changes relevant to HIV transmission."[220] While the authors acknowledged that the most direct measure of the prevention value of needle exchange is the extent to which HIV prevalence and seroconversion rates changed, they nonetheless chose a survey methodology to circumvent several practical and logistical constraints. These constraints included: (1) a brief evaluation period that was insufficient for measuring seroconversion rates; (2) anticipated reluctance of persons to participate in a study requiring blood testing; and (3) a reluctance to engage in invasive investigative techniques within the context of "an innovative service establishing its credentials among a cautious population."[221]

---

contact with methadone programme during the previous 5 years, [and] having volunteered for an HIV test." *Id.*

[215] Dependent variables for activity risk levels were measured as follows: 0 for no injection in the previous month; 1 for not borrowing needles and using only sterile needles; 2 for not borrowing needles but re-using one's own needles; and 3 for borrowing used needles. *Id.*

[216] *Id.* at 573.

[217] *Id.*

[218] *Id.* at 575.

[219] Martin C. Donoghoe et al., *Changes in HIV Risk Behaviour in Clients of Syringe-Exchange Schemes in England and Scotland*, 3 AIDS 267 (1989).

[220] *Id.* at 268.

[221] *Id.*

While the investigators compiled basic client profiles during each participant's first visit to the exchange, their analysis focused on two sequentially administered client questionnaires called CQ1 and CQ2, which were conducted subsequent to the initial visit. CQ1 collected baseline measures of attitudes, knowledge, and behaviors within the first month of attendance. CQ2 collected the same measures two to four months later.[222] While the sample studies comprised only six percent of all clients who had attended the exchanges once over a one-year period, comparative data from the intake sheets of all clients suggested that the sample was representative of the exchange population in regard to both reason for participation and percentage engaging in syringe-sharing behavior at intake.[223] Differences between sample and population were observed in gender, age, age at first injection, and treatment history.[224]

The authors openly acknowledge the methodological limitations of their study. Because the needle exchange programs were open to all persons who wished to enroll, the population at the exchanges was not a randomized sample of injection drug users.[225] The authors also observe that the survey respondents "could be affected by other factors: they were potentially exposed to the government anti-injecting media campaign and the widespread coverage of AIDS and drug use in the press and on television."[226] Subject to these limitations, they observed from CQ1 to CQ2 a reduction in the number of respondents who had engaged in needle sharing or borrowed needles within the previous four weeks.[227] The authors point out several trends in this regard, including a decrease from thirty-four percent to twenty-seven percent on the measure of "whether or not the person shared in the previous four weeks," a change from twenty-five percent to nineteen percent "in the proportion injecting with equipment that had been used by someone else," and a decrease from thirty percent to twenty-five percent "in the proportion passing on used equipment."[228]

---

[222] *Id.*
[223] *Id.*
[224] *Id.*
[225] *Id.*
[226] *Id.*
[227] *Id.* at 269.
[228] *Id.*

The authors also observed syringe-sharing behaviors among drug users not participating in needle exchange programs at the time of the surveys. They note that "[a]mong non-attenders, syringe-sharing in the previous four weeks was reported by 62% at Time 1 and 59% at Time 2."[229] When asked why they engaged in needle sharing activity, the proportion of respondents that cited difficulty in obtaining equipment fell dramatically, from fifty-four percent in CQ1 to nine percent in CQ2.[230] Likewise, the percentage of respondents that gave cost of obtaining apparatus as a reason for sharing declined from sixteen percent in CQ1 to two percent in CQ2.[231]

The authors conclude that their evidence "suggests that people who continued attending syringe-exchange schemes reported small but important reductions in HIV risk behavior in the short-term."[232] While these changes are encouraging, a number of questions remain to be addressed: (1) To what extent do the methodological limitations observed by the authors undermine the small-scale changes observed? In particular, the fact that syringe sharing behavior fell even among the comparison group of non-exchange participants suggests that behavioral changes might indeed be partially attributable to extrinsic variables, such as media coverage and government campaigns; (2) Are short-term findings sufficient to govern policy decisions? If the low-level changes observed in the short-term cannot be sustained in the long run, the ultimate effect of short-term behaviors on seroconversion rates may be relatively small; (3) Do these findings raise policy issues regarding the optimal allocation of scarce resources? Given competing claims on public dollars, including but not limited to the claims of other HIV and AIDS prevention programs, financial support for needle programs may not be the best use of funds.

Survey research, even under the most propitious circumstances, is subject to errors in information reporting.[233] Incentives to lie or exaggerate are substantial when the behaviors being studied

---

[229] *Id.* at 271.
[230] *Id.* at 269.
[231] *Id.*
[232] *Id.* at 271.
[233] These errors fall within two kinds of categories—reporting inaccuracies in which (1) the respondent is attempting to answer honestly but fails to do so because of self-deception, and (2) the respondent knowingly reports false information to deceive the surveyor.

are illegal,[234] stigmatized,[235] or both.[236] Moreover, surveys taken among injection drug users are susceptible to inaccuracies associated with cognitive impairment due to drug usage.[237] Despite these potential hazards, survey research can provide useful information if care is taken to minimize error.[238] This is especially true when the potential sources of error, such as deception of self and others, are unlikely to affect differentially the relevant comparison points, such as needle sharing rates before and after participation in an exchange.

C. *Survey Research in Tandem with Seroprevalence Measures*

Some researchers have combined survey data analysis with examination of seropositivity rates among survey respondents. For example, a study of the Dublin needle exchange by Johnson et al. analyzed data from an anonymous survey, taken in tandem with unlinked, anonymous saliva testing for HIV.[239] The authors found that "60% of HIV positive respondents and 34% of HIV negative respondents had recently[240] shared equipment with two or more partners, despite freely available clean works and intensive counselling . . . ."[241]

The authors temper a "serious concern" over this high level of insider sharing with the observation that "it is quite possible that an even higher level of needle-sharing would prevail if the ex-

---

[234] For a discussion of various illegal aspects of injection drug use, see *supra* notes 26–28 and accompanying text.

[235] *See* JAMES L. SORENSEN ET AL., PREVENTING AIDS IN DRUG USERS AND THEIR SEXUAL PARTNERS 168 (1991) (referring to stigmatization of injection drug users).

[236] *Id.* Under this reasoning, reporting error may increase in cultural contexts which adopt legal sanctions against drug use, rather than the less threatening harm reduction approaches mentioned earlier.

[237] *See* T. Lampinen et al., *Intravenous Drug Users and Human Immunodeficiency Virus Testing and Counseling*, 262 JAMA 1331 (1989).

[238] Those who voluntarily agree to participate in needle programs evince at least some degree of trust in the process, so the degree of calculated deception likely to impair survey research accuracy under a variety of cultural conditions should be studied rather than presumed. Together with studies indicating true and erroneous survey response patterns, investigators may be able to adjust data gleaned from interviews to achieve increased accuracy. Even accounting for survey response error, social scientists generally consider the methods useful and important. For a general discussion of survey research methodology and its uses and benefits, see EARL R. BABBIE, SURVEY RESEARCH METHODS (2d ed. 1990).

[239] Z. Johnson et al., *Prevalence of HIV and Associated Risk Behaviour in Attendees at a Dublin Needle Exchange*, 89 ADDICTION 603 (1994).

[240] The survey asked respondents whether they had shared needles during the month prior to the survey. *Id.* at 604.

[241] *Id.* at 606.

change did not exist."[242] As the authors themselves suggest, without comparative data from which to derive some meaning, their findings provide limited useful information. To interpret the numbers, knowledge about local needle sharing practices outside of the Dublin exchange is necessary. For example, if local outsiders[243] were to engage in levels of needle sharing practices statistically indistinguishable from those of insiders, the exchange would appear to have no relationship to short-term transmission rates. If outsiders were to engage in significantly greater sharing practices than insiders, however, the needle exchange might arguably have some short-term benefit.[244] Without comparative data, investigators are left with nothing more than conjecture regarding their hypotheses.

In 1987, Wolk et al. employed analysis of questionnaire data and testing of returned syringes for presence of HIV to examine two needle exchange programs in Sydney, Australia.[245] The investigators collected used syringes every week from two inner-city exchange sites, and sent them to a laboratory where a random sample of the collected syringes was tested for the presence of HIV antibodies.[246] They noted that this method is likely to provide more accurate information regarding seroprevalence rates among injection drug users than previous studies that analyzed blood samples taken from drug center treatment participants.[247] Because HIV seroprevalence among injection drug users in treatment tends to be lower than among those not in treatment, testing syringe samples from a nontreatment source provides a more inclusive, and therefore potentially more accurate, tally of overall seroprevalence levels.[248]

---

[242] *Id.*

[243] "Outsiders" here refers to those injection drug users in the relevant population—in this case, in Dublin—who do not participate in the needle exchange, and serve as an imperfect form of nonexperimental control. "Insiders" refers to those injection drug users who participate in the needle exchange being studied.

[244] Even if needle sharing rates among insiders are found to be lower than among outsiders, investigators should carefully consider the need to eliminate variables other than insider versus outsider status that might account for any variance observed between the two groups. For example, if insiders volunteer to participate in the needle exchange, they may be inherently more responsible or inherently more concerned with their health than those who decline to participate.

[245] Jael Wolk et al., *Syringe HIV Seroprevalence and Behavioral and Demographic Characteristics of Intravenous Drug Users in Sydney, Australia, 1987*, 2 AIDS 373 (1988).

[246] *Id.* at 374.

[247] *Id.* at 375.

[248] *Id.*

The investigators analyzed syringes derived from the two exchange locations separately, testing twice at the Exchange 1 location to note any change in seropositivity rates over a seven-month period. At Exchange 1, syringe seroprevalence, or percentage of tested syringes found positive for HIV antibodies, was one percent the first time, and 1.5% seven months later—a difference not statistically significant at a ninety-five percent confidence level.[249] At Exchange 2, where the syringes were tested only once, the seroprevalence rate was six percent, significantly higher than at Exchange 1.[250]

In their discussion, the investigators observed, "[t]he finding that there was no statistically significant increase in the syringe HIV seroprevalence at Exchange 1 supports the view that the availability of sterile syringes may have prevented transmission of HIV among the clients of this exchange."[251] While this statement is technically true by virtue of the authors' use of the phrase "may have prevented," it is not a very telling statement, nor are the data here particularly revealing. The lack of change may indicate most anything, and the data provide us with no interpretive aids. Given that stasis may indicate the ineffectiveness of an independent variable,[252] we cannot reasonably attribute seroprevalence rate stability, even cautiously or speculatively, to needle exchange intervention.

This kind of study is most helpful when the authors limit the discussion of their findings to reasonable inferences. For example, the authors note that an overall syringe HIV seroprevalence of three percent for the two exchanges combined is significantly higher than the one percent rate observed in 1986.[253] They also observe substantial needle sharing activity over a six month period,[254] which suggests that the problem of needle-borne transmissions in Sydney is probably increasing in magnitude. These observations are valuable in their ability to provide accurate, comprehensive information regarding seroprevalence among in-

---

[249] *Id.* at 374.

[250] *Id.*

[251] *Id.* at 375.

[252] Indeed, although the reasoning is potentially flawed, we generally presume that lack of change in a dependent variable indicates lack of effectiveness of independent variables. While a change may always occur in the absence of some intervening effect, we should not presume this to be the case unless we have a reason to expect change to be the baseline or default assumption in the absence of the influence or force being examined.

[253] Wolk et al., *supra* note 245, at 375.

[254] *Id.*

jection drug users in the city. Such information may be less ambitious than inferences regarding the potential effectiveness of needle programs in curbing the spread of AIDS, but it does have policy implications, particularly in regard to resource allocation decisions in governmental budgeting processes.

D. *Seroconversion Analyses that Employ Multivariate ANOVA*

Among the most ambitious studies that measure seroconversion rates are Kaplan and Heimer's tests of the "circulation theory" of needle exchange efficacy, which apply multivariate analysis of variance techniques to help explain observed effects. According to this theory, "as clients visit the exchange more frequently, the time needles remain in circulation decreases . . . . As a consequence, needles have a lower probability of becoming infected, and those sharing needles have a lower risk of infection."[255]

Kaplan and Heimer have focused their research in this area on New Haven's needle exchange program. The authors observe that studies that entail periodic HIV testing of participants provide useful information concerning new transmissions and the validity of the circulation theory, but can be logistically challenging.[256] This is especially true in the United States, where criminalization and stigmatization of drug use foster participant mistrust of official agencies and undermine rates of regular participant return over the life of a longitudinal study.[257] Given these concerns, it is not surprising that the New Haven needle program studied by Kaplan and Heimer engaged in neither mandatory nor voluntary sequential HIV testing of participants, in an effort to avoid processes likely to dissuade otherwise willing participants.[258]

---

[255] Edward H. Kaplan et al., *A Decline in HIV-Infected Needles Returned to New Haven's Needle Exchange Program: Client Shift or Needle Exchange?*, 84 AM. J. PUB. HEALTH 1991, 1991 (1994).

[256] Edward H. Kaplan & Robert Heimer, *HIV Incidence Among Needle Exchange Participants: Estimates from Syringe Tracking and Testing Data*, 7 J. AIDS 182, 182 (1994) ("To measure HIV *incidence* among needle exchange clients directly would require sequential testing of participating drug injectors over time; such studies would in theory provide the best measures of incidence. The paucity of long-term incidence studies in needle exchanges or other street outreach settings testifies to the difficulty of maintaining serial HIV testing in such populations . . . .").

[257] Rates of regular participant return should decline as a function of two factors: (1) fear of criminal or other official sanctions, and (2) the inability of incarcerated drug users to continue participation.

[258] Kaplan & Heimer, *supra* note 256, at 183.

Kaplan's work with Heimer and other colleagues was an ambitious attempt to circumvent these kinds of logistical limitations, as well as methodological limitations we have already observed. The New Haven program tracked syringes returned to the exchange and tested them for HIV-1 proviral DNA prevalence.[259] Heimer et al. observed a DNA prevalence of 67.5% at the start of the program,[260] followed by a rapid decline to below forty-five percent within the first three months. DNA prevalence remained stable at this level for the next ten months.[261] The authors observed no meaningful changes in either the demographics or the drug use habits of participants, and attributed the decrease in tainted syringes to the effectiveness of the needle program.[262]

The New Haven data yielded several reports of early findings that consistently indicated a decrease in the presence of HIV in syringes returned to the exchange over time.[263] The analysis of data from which the encouraging findings emerged required "a number of untested behavioral assumptions regarding frequency of drug injection, frequency of sharing, frequency of needle cleaning, and the ways in which [intravenous drug users] interact when sharing needles," as well as some reliance upon self-reported data.[264] Moreover, the results were criticized by the Office of National Drug Control Policy on the grounds that the observed reduction in HIV prevalence in the studies could be a function of a "temporal shift in demographics of persons using the needle exchange."[265] This alternative interpretation of the decline, known as "the client-shift hypothesis," might be valid if, for example, the program "initially attracted HIV seropositive and higher-risk seronegative injection drug users, but over time,

---

[259] Robert Heimer et al., *Needle Exchange Decreases the Prevalence of HIV-1 Proviral DNA in Returned Syringes in New Haven, Connecticut*, 95 AM. J. MED. 214 (1993).

[260] *Id.* at 216.

[261] *Id.* at 214, 217.

[262] *Id.* at 218–20.

[263] Over the course of the period studied, the researchers observed a reduction in indicators of HIV of about 33%. *See also* Edward H. Kaplan & Robert Heimer, *HIV Prevalence Among Intravenous Drug Users: Model-Based Estimates from New Haven's Legal Needle Exchange*, 5 J. AIDS 163 (1992); Edward H. Kaplan & R. O'Keefe, *Let the Needles Do the Talking! Evaluating the New Haven Needle Exchange*, 23 INTERFACES 7 (1993).

[264] Kaplan & Heimer, *supra* note 256, at 183.

[265] David Vlahov & Ronald S. Brookmeyer, *Editorial: The Evaluation of Needle Exchange Programs*, 84 AM. J. PUB. HEALTH 1889, 1890 (1994) (citation omitted).

enrollment . . . extended to lower-risk injection drug users who were more likely to be and to remain HIV seronegative."[266]

Kaplan and Heimer's follow-up analyses sought to resolve these shortcomings. The investigators employed a "maximum likelihood change point model" that infers likelihood of infection solely from needle analysis, without reliance upon any assumptions concerning rates of injection, needle cleaning, or needle sharing.[267] Syringe tracking and testing yielded individual time series data, which were subject to sophisticated analytical techniques designed to determine the statistical likelihood that any individual became infected.[268] The authors contend that the techniques permitted accurate estimation of HIV incidence among needle exchange participants, and that replacement of self-reporting and sequential testing of participants with syringe tracking and testing data yielded accurate information without undue intrusiveness.[269] The authors found that their "best estimate for the number of new infections that occurred during the first 19 months of the New Haven needle exchange [was] zero."[270]

To support the "circulation theory" against alternative "client-shift" hypotheses,[271] the investigators employed multivariate analysis of variance (MANOVA), a process which employs statistical regression techniques to identify the degree to which a number of variables predict an observed effect[272]—in this case, the decline in needle seropositivity. The MANOVA revealed that "mean needle circulation time significantly predicted the level of HIV-positive needles, with increasing circulation time corresponding to increasing HIV prevalence in needles."[273] Furthermore, only one of nine demographic variables examined[274]—race—emerged

---

[266] *Id.*

[267] Kaplan & Heimer, *supra* note 256, at 184.

[268] *Id.*

[269] *Id.* at 187.

[270] *Id.*

[271] *See supra* note 265 and accompanying text.

[272] Analysis of variance techniques explain the degree to which any of a number of variables predict a given phenomenon. Inference of causality cannot be presumed without the employment of appropriate methodological techniques. For a brief description of the uses and limits of analysis of variance, see JOHN E. FREUND & GARY A. SIMON, MODERN ELEMENTARY STATISTICS 393–423 (8th ed. 1992).

[273] Kaplan et al., *supra* note 255, at 1993.

[274] The nine variables analyzed in the MANOVA, chosen "because of their association with an increased likelihood of HIV infection among injection drug users in the northeastern United States," were "sex of client, age at enrollment, duration of drug injection, daily frequency of injection, fraction of time injecting in shooting galleries, fraction of time injecting with shared equipment, fraction of time cleaning injection equipment, injected cocaine (yes or no), and race (White or non-white)." *Id.*

from the MANOVA as having changed significantly over the period in which the reduction in syringe seropositivity was observed. The investigators were able to eliminate even race as a factor in seropositivity reduction, as a logistical regression analysis revealed no significant difference in HIV presence in syringes as coded for race.[275]

These findings provide impressive support for the circulation theory. They suggest that the needle exchange process, and not a shift in needle exchange client demographics over time, decreases new HIV transmission opportunities, and that the more frequent and conveniently available the exchange, the greater the effect. The New Haven investigators view their results cautiously but optimistically, summarizing their findings as "important, additional evidence supporting the efficacy of needle exchange as an HIV prevention program."[276]

E. *Summary*

Examination of these major studies of needle exchange programs to date yields two important observations: (1) needle exchanges appear to have promise as an AIDS prevention tool, but (2) the research base in toto has too many gaps and flaws to be considered at all conclusive. As we shall see in Part V, research findings regarding the effects of needle exchange programs on illegal drug use rates are even scantier. The public policy recommendations in Part VI are based on both the promise and the ultimate inadequacy of the studies we have reviewed.

IV. The Nature and Characteristics of Needle Programs that Appear Most Likely to be Effective

As in any complex issue, the value of needle programs is unlikely to be unequivocally and unconditionally proven or disproven in general, abstract terms. More likely, needle exchanges will produce varying benefits or costs under different conditions.[277] In adopting provisional, experimental needle programs,

---

[275] *Id.* at 1992.

[276] Kaplan & Heimer, *supra* note 256, at 187.

[277] Relative needle exchange success is likely to be contingent upon intelligent and informed choice among many design and implementation variables. A simple example can be posited hypothetically: It is reasonable to believe that some degree of education

policymakers should examine the characteristics of such programs that appear to be effective in curtailing the spread of HIV in various settings. Likewise, should generic pilot studies suggest that needle programs can reduce HIV infection, more refined follow-up studies should be designed to provide information useful in structuring more promising programs for the future. While studies have already begun to reveal some characteristics of successful programs, the literature is incomplete. The research discussed in this subsection should be supplemented with further study aimed at helping policymakers design the best possible pilot programs.[278] What follow are some observations gleaned from the research to date which suggest the qualities of a promising needle program.

A. *Frequency and Convenience of Operations*

Frequency and convenience of operations are likely to enhance needle program effectiveness in reducing HIV transmission opportunities. As observed in the preceding part, needle exchange effectiveness is associated with what Kaplan et al. call the circulation theory: frequent visits to needle exchanges should abbreviate needle lifespans and reduce circulations per needle, thereby diminishing infection opportunities.[279] The critical behavioral changes sought under needle programs are the elimination or reduction of needle sharing, both of which are supported by an exchange's maintenance of regular, convenient hours of operation. In creating a needle exchange and organizing its logistical and operational details, designers should carefully con-

---

will be associated with an effective needle exchange, so that a maximum number of exchange users know how to exploit potential exchange-related risk reduction strategies effectively. Investigators can learn much about needle exchange effectiveness by testing a variety of such potential factors and the impact they have on relevant dependent variables.

[278] Policymakers and investigators should apply the limited resources available to create programs that promise the greatest chance of achieving encouraging results. Were needle programs purely scientific endeavors, it might be reasonable and defensible to test all types of program possibilities. When investigation is gleaned in field studies that directly affect the lives of subjects, the public service component of the research agenda creates ethical concerns. It is reasonable under these circumstances to limit pilot projects to the most promising prospects. Such projects will include those most likely to save lives and uncover information leading to effective, workable needle programs for the future. Investigators should remain open to the possibility that even the most promising types of needle programs may not survive cost-benefit analysis or may engender bad public policy within certain environments, contexts, or cultures.

[279] Kaplan et al., *supra* note 255.

sider access and convenience.[280] Recommendations that might impair availability of needles and syringes should be rejected if they fail to serve sufficiently important countervailing goals.

For example, some proponents of urban redevelopment have recommended that cities rotate the location of needle exchange programs to discourage the degradation of program neighborhoods.[281] This recommendation should be carefully scrutinized prior to implementation. Rotation of location is likely to create confusion and misunderstanding among exchange users, resulting in a reduction in consistent utilization rates.[282] The goal of rotation—improvement of the safety and accessibility of neighborhoods[283]—should be evaluated on three levels: whether the goal is realistic, whether the recommended rotation is likely to achieve the goal, and whether the magnitude of projected benefits outweighs any anticipated cost to the program.[284]

Likewise, regulations that would employ pharmacists to execute needle exchange functions must be carefully evaluated. As needle exchange programs frequently require an exception to[285] or the circumvention of[286] state drug paraphernalia laws, ques-

---

[280] It is critical to consider the optimal degree of accessibility and convenience of needle exchange program hours and locations. Given the scarcity of resources, economic analysis of the cost-benefit tradeoffs at various levels of accessibility and convenience can provide lawmakers with extremely important information. Such studies can help identify the incremental costs associated with marginal accessibility and convenience, as well as the points at which marginal returns diminish.

[281] *See* Sylvia W. Nogaki, *Forum Seeks Ways to Make Downtown Corridor Friendlier—Sprucing Up Pine Street*, SEATTLE TIMES, Mar. 25, 1995, at C1 (citing recommendation of Thomas Harville that Seattle consider rotation to maintain the attractiveness of the neighborhood).

[282] Indeed, Guydish et al. found higher needle return rates for fixed exchange sites than for mobile exchange sites. *See* Joseph Guydish et al., *supra* note 175 at 617–19.

[283] Concern about a site for a permanent needle exchange or the stopping point of a mobile needle exchange frequently begins with protests over a proposed needle exchange in the protestor's home neighborhood. *See, e.g.*, Kevin Diaz, *How to Really Needle Someone: A Radio Host and a Businessman Get into a (Publicity?) Exchange*, STAR TRIB. (Minneapolis), Mar. 25, 1995, at 1B; Barbara Clements, *East Side Says Needle Exchange Unwelcome There*, NEWS TRIB. (Tacoma, Wash.), Mar. 23, 1995, at B1.

[284] The goal may be unrealistic if, for example, rotation of needle exchange sites only serves to shift problems to other neighborhoods without improving overall safety levels. In addition, rotation might yield only nominally improved safety conditions that are outweighed by benefits of exchange accessibility that accompany stability of location.

[285] *See supra* note 33 and accompanying text.

[286] Some cities in California have attempted to circumvent state laws by declaring a state of emergency. *See supra* note 37 and accompanying text. The San Francisco Board of Supervisors has recently extended such a resolution. *See* Clarence Johnson, *Boardwatch*, S.F. CHRON., Apr. 25, 1995, at A14 (reporting approval by a vote of 9 to 0 of a motion to extend a declaration of an emergency in order to continue the needle exchange program).

tions have arisen regarding the appropriate role of pharmacists in needle exchange programs.[287] Along these lines, the CDC has recommended the repeal of state syringe prescription paraphernalia laws; the encouragement by local communities of syringe sale, distribution, or exchange by pharmacists; and surveys of pharmacists regarding their "willingness to participate in pharmacy-based syringe sale, distribution, or exchange . . . ."[288]

Any proposed use of pharmacists to distribute needles should be examined to consider effects on needle distribution levels. Distribution of needles and syringes within pharmacies may increase sterile needle use by expanding the number of available exchange sites and by adding a setting that may be less intimidating to some than existing clinical sites. Conversely, the use of professionals as required intermediaries in existing street exchanges runs the risk of constraining operations by limiting the number of qualified exchange personnel and frightening off potential program participants more comfortable dealing with nonprofessionals or outreach workers[289] who have developed a presence and trust in the community.[290]

Exchange site rotation and pharmacist utilization decisions share a characteristic common to all choices regarding the design of optimal experimental needle programs: they should consistently reflect the underlying logic behind needle exchange, i.e., that frequent and exclusive use of the facility will reduce the life of a needle and the number of injectors who use it.[291] Qualities that enhance an exchange's accessibility encourage frequent, exclusive use, thereby improving the potential utility of the program.

---

[287] William A. Zellmer, *Pharmacist Involvement in Needle Exchange Programs*, AMER. PHARMACY, Sept. 1994, at 48 (citing P. LURIE & A.L. REINGOLD, THE PUBLIC HEALTH IMPACT OF NEEDLE EXCHANGE PROGRAMS IN THE UNITED STATES AND ABROAD (1993)).

[288] *Id.* at 49.

[289] For discussion of the role of outreach workers who are former drug users, indigenous to communities served, see Claire Sterk-Elifson, *Outreach among Drug Users: Combining the Role of Ethnographic Field Assistant and Health Educator*, 52 HUM. ORG. 162 (1993).

[290] Trust is an important consideration when people are asked to expose themselves as drug users to anyone who appears to act in some official capacity. Part of the accessibility of needle exchange programs to drug users will be the provision of apparent safety and security from any criminal law enforcement. This may explain in part the findings of Paone et al., in which significantly more syringes were returned to indoor sites than to outdoor sites. Denise Paone et al., *Operational Issues in Syringe Exchanges: The New York City Tagging Alternative Study*, 20 J. COMM. HEALTH 111, 121 (1995).

[291] Kaplan et al., *supra* note 255.

B. *Utilization of a Counseling Component*

Needle program effectiveness appears to be enhanced when a well-formulated, well-implemented counseling component[292] is included.[293] Any success of needle exchanges in the reduction of HIV transmissions will be directly linked to the elimination of needle sharing practices. Unfortunately, providing clean needles in the absence of counseling may fail to produce consistent behavioral changes.[294] Untrained program participants can commit a variety of errors based on faulty understanding of how HIV is transmitted and how the risk of transmission is diminished.[295] For example, some participants might be susceptible to a misunderstanding and misapplication of analogues from public health. A user who correctly believes that a reduction in smoking or drinking will tend toward better health may believe that a reduction in needle sharing behaviors is a reasonable compromise between exclusive exchange use and needle sharing. This speculation is consistent with research findings that associate educational intervention programs with short-term reductions in both injection drug use and syringe sharing.[296]

---

[292] AIDS counseling and education are treated collectively here, under the assumption that counseling can serve educational functions that help retard transmission of HIV among injection drug users. For discussion of HIV counseling and education, see generally Sari H. Dworkin & Lester Pincu, *Counseling in the Era of AIDS*, 71 J. COUNS. & DEV. 275 (1993); Reese M. House & Catherine M. Walker, *Preventing AIDS via Education*, 71 J. COUNS. & DEV. 282 (1993); Richard P. Keeling, *Commentary: Educating and Counseling about HIV in the Second Decade*, 71 J. COUNS. & DEV. 306 (1993).

[293] Van den Hoek et al., *supra* note 198, at 1357. For detailed discussion of the nature of effective counseling programs as recommended by various investigators and commentators, see Willard Cates, Jr. & H. Hunter Handsfield, *HIV Counseling and Testing: Does it Work?*, 78 AM. J. PUB. HEALTH 1533 (1988); Wallace Mandell et al., *Changes in HIV Risk Behaviors Among Counseled Injecting Drug Users*, 24 J. DRUG ISSUES 555 (1994).

[294] The development of consistent behavioral changes requires an understanding of the means by which HIV is transmitted and the increased risk of transmission associated with even occasional needle sharing. *See* Lindsay Kines, *MD Asks Premier to Help Curb HIV: Virus "Spreading Like Wildfire" Among Injection Drug Users*, VANCOUVER SUN, May 12, 1995, at B7 (citing a medical clinic director's belief that needle exchanges must be supported by education and counseling to alter high-risk behaviors).

[295] *But see* Donald A. Calsyn et al., *Ineffectiveness of AIDS Education and HIV Antibody Testing in Reducing High-Risk Behaviors Among Injection Drug Users*, 82 AM. J. PUB. HEALTH 573, 575 (1992) (suggesting a high level of knowledge among injection drug users regarding HIV and the means of its transmission and noting that alternative interventions therefore may be more effective than information-giving approaches).

[296] Richard C. Stephens et al., *Effects of an Intervention Program on AIDS-Related*

The exclusive use of needles that have been distributed through an exchange may be facilitated by another aspect of counseling: communication of planning strategies that may help addicts avoid the desperate situations and concomitant carelessness that can accompany addiction.[297] Counseling can serve not only to teach these strategies, but also to provide the social reinforcement and support that help participants maintain their commitment to the principles of the exchange.[298]

Substantial, high-quality counseling programs associated with needle exchanges can also assist in the referral of drug users to treatment facilities.[299] To the degree that needle programs can serve this secondary referral role, they might reduce rather than increase injection drug use rates,[300] thereby allaying fears that needle exchange will undermine efforts to reduce drug usage.[301] Any net reductions in illegal drug use rates are valuable not only in themselves, but also in eliminating some HIV transmission opportunities.

Finally, counseling services provided by needle exchange personnel have the potential to lower intravenous transmission rates by furnishing information or risk reduction tools associated with other means of transmission.[302] For example, if on-site counseling regarding safer sex practices or on-site distribution of condoms were to reduce HIV prevalence among injection drug users, the total number of transmission opportunities between needle sharers logically would diminish. Given evidence of a

---

Drug and Needle Behavior Among Intravenous Drug Users, 81 AM. J. PUB. HEALTH 568 (1991).

[297] These kinds of strategies fall within the broader context of what has been identified as "skills-building intervention" for the reduction of HIV transmission among methadone patients. For discussion of this skills building approach, see Robert F. Schilling et al., *Skills-Training Groups to Reduce HIV Transmission and Drug Use Among Methadone Patients*, 40 SOC. WK. 91 (1995).

[298] Effective needle programs appear to require some form of social reinforcement and support. As one commentator has observed, changes in needle sharing behaviors through needle program utilization require that the participant be motivated and supported to engage in healthier behaviors. Lawrence W. Green, *Does Needle-Exchange Save Lives?*, AM. J. SURGERY, Sept. 1993, at I, II.

[299] Alan A. Wartenberg, *'Into Whatever House I Enter': HIV and Injecting Drug Use*, 271 JAMA 151 (1994).

[300] While it seems reasonable to speculate that drug treatment referrals might prevent some HIV infections as successful participants stop injecting drugs, this hypothesized trend should be studied. The magnitude of any trend should be measured against any countervailing negative effects with respect to drug use rates.

[301] *See supra* notes 140–142, 149–152 and accompanying text. For further discussion of the question of the relationship between needle exchange programs and drug use rates, see *infra* part V.

[302] Wartenberg, *supra* note 299, at 151.

relationship between drug use and high-risk sexual activity,[303] an opportunity to address both issues in tandem is a promising component of needle exchange counseling services.

For each of the counseling functions discussed in this subsection, some variants in counseling design will prove more effective than others. Studies assessing the comparative efficacy of assorted interventions have already begun to shed light on the optimal design of educational and counseling services under a variety of conditions.[304] Designers should assess this work in developing counseling programs that optimally complement experimental needle exchange programs.

C. *Customization of Needle Program Design*

Needle programs should be tailored to the characteristics of their target populations. The programs serve constituencies that function within cultural contexts.[305] The better we understand these constituencies and their underlying social networks,[306] the better we can match program development to user characteristics[307] and the more effective a program is likely to be.[308] This subsection addresses research that helps us understand target populations.[309] The program developers can use such studies to shape specific program qualities.

---

[303] For discussion of studies in this area, see Barbara C. Leigh & Ron Stall, *Substance Use and Risky Sexual Behavior for Exposure to HIV: Issues in Methodology, Interpretation, and Prevention*, 48 AM. PSYCH. 1035 (1993).

[304] *See, e.g.*, Jane McCusker et al., *AIDS Education for Drug Abusers: Evaluation of Short-term Effectiveness*, 82 AM. J. PUB. HEALTH 533 (1992).

[305] For discussion of social and cultural contexts of HIV and AIDS programs in general, see James M. Croteau et al., *Social and Cultural Sensitivity in Group-Specific HIV and AIDS Programming*, 71 J. COUNS. & DEV. 290 (1993).

[306] Neaigus et al. have distinguished two important kinds of networks that are relevant to prevention of HIV infection: risk networks and social networks. Risk networks refer to groups within which risk behaviors serve as vectors of the transmission of disease; social networks refer to the generation and dissemination of social influence. Alan Neaigus et al., *The Relevance of Drug Injectors' Social and Risk Networks for Understanding and Preventing HIV Infection*, 38 SOC. SCI. & MED. 67 (1994).

[307] *See* David R. Holtgrave et al., *An Overview of the Effectiveness and Efficiency of HIV Prevention Programs*, 110 PUB. HEALTH REP. 134, 135 (1995) (suggesting tailoring HIV prevention messages to the characteristics of the audience).

[308] Adapting program design to prospective users should permit programs to reach a larger group within a shorter time frame. This effect is an essential component of "impact effectiveness" and therefore of the potential to alter the course of the HIV epidemic. Jeff Stryker et al., *Prevention of HIV Infection: Looking Back, Looking Ahead*, 273 JAMA 1143, 1145 (1995).

[309] This emphasis on understanding target populations and tailoring programs to those populations is not meant to imply a recommendation that prevention programs be

In a review of the literature, Hartgers et al. cite numerous factors found to be related to needle and syringe sharing behavior: "multiple drug use, younger age, homelessness, cocaine use (including injecting), injecting drug use by a regular partner or by peers, drug craving, little experience with injecting, and frequency of injecting."[310] In their own study of 131 seronegative injecting drug users in Amsterdam, Hartgers et al. identified a similar group of specific factors associated with high risk of needle and syringe borrowing: long-term, moderate to heavy alcohol use; current engagement in cocaine injections; and lack of permanent housing.[311]

The authors of the study suggest that effective needle programs should be tailored to the needs of those most likely to share needles.[312] Because cocaine injectors, long-term alcohol users, and the homeless all may experience difficulty in long-term planning, mitigation of needle borrowing within this population set may depend upon programs that ensure immediate access.[313] For example, extended hours of operation and increased numbers of needle exchange sites may be important factors in reducing borrowing among irregular needle exchangers.[314]

Whereas the study by Hartgers et al. focuses on behavioral characteristics of needle sharers, Sherry L. Elnitsky and Thomas J. Abernathy also examined the demographic characteristics of needle exchange users.[315] The study of 103 participants who volunteered to be interviewed (eleven percent of estimated exchange users)[316] found a male-to-female ratio of three to one and

---

targeted exclusively or predominantly to so-called "marginalized" groups. The issue of whether AIDS prevention should be targeted or mass-marketed goes beyond the analysis in this Article. Instead, the observations here simply suggest that needle programs are by nature geographically constrained entities, and as such may benefit from an understanding of local cultures and demographics. For discussion of the controversy over whether to treat AIDS as a nationalized or a local problem, see Don C. Des Jarlais et al., *Targeted HIV-Prevention Programs*, 331 N. ENG. J. MED. 1451 (1994); David E. Rogers & June E. Osborn, *AIDS Policy: Two Divisive Issues*, 270 JAMA 494 (1993).

[310] Christina Hartgers et al., *Needle Sharing and Participation in the Amsterdam Syringe Exchange Program Among HIV-Seronegative Injecting Drug Users*, 107 PUB. HEALTH REP. 675 (1992).

[311] *Id.* at 679.

[312] *Id.* at 681.

[313] *Id.*

[314] *Id.*

[315] *See* Sherry L. Elnitsky & Thomas J. Abernathy, *Calgary's Needle Exchange Program: Profile of Injection Drug Users*, 84 CAN. J. PUB. HEALTH 177, 177 (1993). Elnitsky and Abernathy studied Calgary's Injection Drug Education and Prevention Program, implemented under Canada's national AIDS strategy.

[316] *Id.*

an average user age of thirty-two, with fifty-eight percent age thirty or older.[317] The authors also analyzed the ethnicity, employment status, and relative permanence or transience of participants, as well as behavioral and cognitive factors such as their age of first injection, knowledge regarding AIDS, history of HIV testing, and other drug use and sexual habits.[318]

While the creation of such demographic profiles of needle exchange users does not tell us whether the drug program in Calgary will effectively reduce future transmissions via shared needles, it does provide information that can be very useful in tandem with effectiveness studies like those discussed in Part III. Knowledge of the demographic characteristics, personal characteristics, and behavioral patterns of participants can provide clues that help designers fine-tune an experimental program.[319] For example, levels of knowledge regarding AIDS and its transmission may affect both the quantity and quality of resources committed to the educational counseling component of an exchange. Likewise, gender demographics may highlight the importance of maintaining daytime hours of operations, particularly in areas where violent crimes against women occur at high rates.[320] Racial and ethnic composition of constituent populations should also be considered in making such fundamental decisions as whether to employ multilingual staff.[321] Likewise, age of ex-

---

[317] *Id.* at 178.

[318] *Id.* at 178–79.

[319] Understanding the demographic and behavioral characteristics of needle exchange users can help policymakers to shape and understand effective programs at two levels—at the hypothesis formulation/program design level and at the hypothesis testing/program monitoring level. Data concerning an exchange's constituency will drive intelligent speculation about how the exchange will work, based on the needs, drives, and limitations of particular user bases. These data can assist program designers in making informed choices based on rational expectations as they create the structures that accompany program organization. Moreover, continued collection of demographic and behavioral data during the life of the exchange can be used to test the assumptions upon which intelligent design decisions were originally made.

[320] For a more general discussion of intervention dynamics for women at high risk for HIV, see Jeffrey A. Kelly et al., *The Effects of HIV/AIDS Intervention Groups for High-Risk Women in Urban Clinics*, 84 AM. J. PUB. HEALTH 1918 (1994).

[321] Needle programs should be sensitive to cultural characteristics of constituent groups. Some racial and ethnic groups are at a comparatively high risk of contracting HIV. Among some groups, both the number of diagnosed AIDS cases and the rate of increase in HIV transmissions are disproportionately high. A CDC report in 1988 noted that 26% of Americans with AIDS were black and 14% were Hispanic, whereas blacks constituted only 12% and Hispanics six percent of the U.S. population. *Centers for Disease Control: AIDS Wkly. Surveillance Rep.*, June 6, 1988; *HIV/AIDS in Racial/Ethnic Communities: The Burden of Disease*, CDC HIV/AIDS PREVENTION NEWSLETTER (Centers for Disease Control, Atlanta, Ga.), Oct. 1990, at 3. In U.S. cities, these phenomena appear to be associated with transmission among drug users. In some areas,

change users may also indicate desirable program characteristics. For example, the relatively advanced age among program users in the Calgary study[322] may suggest specific counseling needs that differ from the counseling needs of younger populations.[323]

D. *Consideration of Creative Variants in Needle Program Design*

If traditional needle exchange programs continue to show promise, creative variants should be examined, including variants that may be politically unpopular. Specifically, options grounded in adherence to the Dutch harm-reduction model[324] should not be dismissed without examination of their effectiveness in Europe and prospects for success in this country. Although the studies discussed in Part III are inconclusive, they do suggest that needle exchanges show promise. Should future refinements in research methodology and design continue to yield favorable results, policymakers should support the tentative adoption of innovative but promising programs, subject to further study for both efficacy and effect on illegal drug usage.

Consider, for example, the concept of community-based needle exchange. Community-based exchange is a form of needle-exchange outreach that has been adopted in some parts of Europe in an effort to expand the reach of successful but more traditional exchanges.[325] The community-based program in Rotterdam was a response to the limited reach of an exchange that was tied to a methadone maintenance program.[326] Because most area heroin users were not enrolled in the maintenance program, and because methadone maintenance was unlikely to attract many

---

blacks and Hispanics constitute the vast majority of injection drug users who have AIDS. Robert F. Schilling et al., *Developing Strategies for AIDS Prevention with Black and Hispanic Drug Users*, 104 PUB. HEALTH REP. 2, 3 (1989). ("In New York City, persons from Hispanic or Afro-American backgrounds account for 86% of IV drug users with AIDS.").

[322] *See supra* text accompanying note 317.

[323] The precise nature of age-contingent needle exchange counseling needs remains to be addressed in the literature. It is reasonable to hypothesize, however, that age affects overall levels of knowledge and skills, and that persons in different age brackets might benefit differently from a variety of drug- and sex-related counseling approaches.

[324] For discussion of the Dutch harm-reduction model, see *infra* note 403 and accompanying text.

[325] For more detailed discussion of community exchange, see Jean-Paul C. Grund et al., *Reaching the Unreached: Targeting Hidden IDU Populations with Clean Needles via Known User Groups*, 24 J. PSYCHOACTIVE DRUGS 41 (1992).

[326] *Id.* at 42.

who switched from heroin to cocaine use, a large segment of the Rotterdam area remained unserved by treatment-related programs.[327]

The collective exchange in Rotterdam is one part of a two-tiered needle program, which employs both traditional individual needle exchange and collective needle exchange.[328] On the collective side of the exchange, outreach workers target pivotal persons who exchange large numbers of needles for numerous users during a single visit.[329] The collective approach allows for exchanges on a massive scale, as used needles are traded for boxes in increments of one hundred.[330] Moreover, by enhancing access,[331] collective processes can broaden the proportion of drug injectors using needle exchange programs. Among those likely to benefit from access to collective exchange rather than individual exchange are the mentally incapacitated, persons physically incapable of personal attendance, persons intimidated by legalistic or bureaucratic contacts regarding illegal drug use, and persons not sufficiently motivated to accept the inconvenience of regular attendance at needle exchanges.

In their comparison of the individual and collective exchange tiers in Rotterdam, Grund et al. found that needle exchange rates were substantially higher among collective exchangers than among individual exchangers—84.8% versus 46.4%.[332] They also found a higher level of exchange use among collective exchangers than among individual exchangers. Whereas collective exchangers constituted only 24% of the exchange's clients, they accounted for 52.6% of the total number of needle exchange contacts.[333] Ethnographic data gathered from the field suggested that "users who engaged in collective exchange were more aware of high-risk

---

[327] The methadone program had contact with about 1000 heroin users, including both smokers and injection drug users. Estimates of heroin users in Rotterdam range from 2500 and 3500. The authors estimate that between 60% and 70% of the target group was not being reached by the needle exchange system tied to methadone maintenance. *Id.* at 42.

[328] *Id.* at 42.

[329] *Id.*

[330] *Id.*

[331] Access to needle exchange programs is best viewed in terms of the number of persons who will use the exchange. For a discussion of the reasons for viewing access to AIDS prevention services in terms of convenience and user friendliness rather than in absolute binary terms of technical accessibility versus technical nonaccessibility, see Steven R. Salbu, *HIV Home Testing and the FDA: The Case for Regulatory Restraint*, 46 HASTINGS L.J. 403, 428–29 (1995).

[332] Grund et al., *supra* note 325, at 44.

[333] *Id.*

behaviors and put more energy into health maintenance and hygiene than individual exchangers."[334]

These findings suggest two possible theories that run counter to conventional principles of effective exchange. First, the higher exchange rate found among collective exchangers may undermine the commonly held belief that "high return rate is the result of strict rules regarding a one-for-one transaction."[335] Given this evidence, further study is needed to confirm or discredit the necessity of one-for-one exchange to accomplish the goal of removing used needles from circulation. If strict exchange rules do not in fact encourage higher exchange rates, they may constrain the potential reach of effective needle programs. Moreover, strict exchange rules may discourage some users from participation.[336]

Second, the ethnographic data findings suggest that professional, on-site counseling of exchange users might not be the optimal mode of effecting behavioral change.[337] Instead, collective exchangers who are also trusted, well-informed insiders may prove to be the most effective counselors. This hypothesis is based on the investigators' observation of attitudes among collective exchange users who evinced both an awareness of risky behaviors and a concern for health maintenance and hygiene.[338] The authors posit that "collective social means" appear to be more effective than "individualistic psychological strategies" in the development of healthy needle exchange habits.[339] They emphasize the role of naturalistic settings and the involvement of drug user networks as potentially effective replacements for what are arguably more artificial on-site staff services.[340] Of course, the ethnographic data are only suggestive and are subject to varying interpretations. Nonetheless, the hypothesis that pivotal fellow drug users can be trained to function as the most effective counselors among other drug users is both plausible and consistent with the study's observations. Should further investigation provide support for the hypothesis, both collective exchanges

---

[334] *Id.*

[335] *Id.* at 45.

[336] Whatever its effect, strict one-for-one exchange may be a necessary component of some experimental programs as a method of tracking needles and inferring information regarding the impact of exchange practices on future HIV transmission rates.

[337] *See supra* text accompanying note 334.

[338] *Id.*

[339] Grund et al., *supra* note 325, at 47.

[340] *See generally* Grund et al., *supra* note 325.

and more organic alternatives to traditional on-site counseling would merit strong consideration. More generally, lawmakers and policymakers should remain open to the examination of unorthodox but promising alternatives to conventional needle exchange programs and should maintain an awareness of those innovative approaches that have yielded favorable results abroad.

V. THE EXTENT, IF ANY, TO WHICH NEEDLE PROGRAMS WILL INCREASE THE INCIDENCE OF ILLEGAL DRUG USE

Critics of needle exchange[341] contend that the programs encourage illegal drug use.[342] They reason that both the appearance of government approval and the provision of injection paraphernalia will exacerbate drug abuse,[343] a social problem that has remained intransigent despite the current war on drugs.[344] In New York, the Chairman of the City Council's Black and Hispanic Caucus recommended the rejection of needle programs, stating that the city sends "the wrong message when it distributes free needles to drug addicts while . . . trying to convince . . . children to say no to drugs."[345] Former Mayor David Dinkins concurred, calling for strict law enforcement policies consistent with the current war on drugs.[346]

---

[341] While the comments in this part apply specifically to needle exchange programs, they have some application to distribution of bleach disinfection kits as well. Distribution of bleach disinfection kits may be somewhat less objectionable to critics than needle exchange on the theory that the government would not be providing the actual paraphernalia used to inject drugs. This distinction may be more technical than meaningful, as both needle and bleach distribution are aimed at helping the drug user in the injection process. For discussion of the bleach disinfection method, see P. Shapshak et al., *Inactivation of Human Immunodeficiency Virus-1 at Short Time Intervals Using Undiluted Bleach*, 6 J. AIDS 218 (1993).

For discussion of limitations of the effectiveness of the bleach disinfection method under field conditions, see David Vlahov et al., *Field Effectiveness of Needle Disinfection Among Injection Drug Users*, 7 J. AIDS 743 (1994).

[342] For a listing of the reasons put forth by critics of needle programs, see *Illicit Drug Use Report*, *supra* note 18, at 946.

[343] *Id.*

[344] *Drug Use Occupies Emergency Rooms, Public Opinion Polls*, 109 PUB. HEALTH REP. 586 (1994).

[345] *Council Calls for End to Free Needles Plan*, N.Y. TIMES, Dec. 7, 1988, at B10.

[346] Todd S. Purdum, *Dinkins to End Needle Plan for Drug Users*, N.Y. TIMES, Feb. 14, 1990, at B1. In accordance with Dinkins' stance against needle exchange, the New York programs were closed briefly in 1990. Underground programs existed in the wake of the ban and were upheld in New York under a necessity defense. People v. Bordowitz, 588 N.Y.S.2d 507 (N.Y. City Crim. Ct. 1991). By 1992, New York City and State authorities were again approving the operations of syringe exchange. For discussion of the history of needle exchange regulation in New York City, see Denise Paone

As noted earlier, critics voice concern about negative repercussions of needle programs using two main arguments—the "public sanction/wrong message" argument and the "provision of tools" argument.[347] The public sanction/wrong message argument suggests that needle programs connote permission to use drugs;[348] the provision of tools argument suggests that needle programs facilitate and encourage drug use by providing users with necessary paraphernalia.[349] Either offhand dismissal or unquestioning acceptance of these assertions is dangerous. Like the issue of needle exchange efficacy in reducing HIV transmissions, the effects of needle programs on illegal drug usage are best addressed by methodologically sound scientific or quasi-scientific investigation.[350] Unfortunately, studies purporting to examine the effects of needle exchange upon rates of illegal drug use have provided little useful information.

The lack of clear, persuasive evidence is a function of several phenomena. The quantity and quality of research regarding program effects on drug usage is low compared to the literature on effectiveness in reducing HIV transmissions among drug injectors.[351] Those investigators who do attempt to analyze program effects upon drug usage usually focus on only one dimension of the issue—whether illegal drug use among needle exchange participants increases, remains stable, or decreases during participation.[352] This kind of study has limited applicability. It does not indicate, for example, whether the existence of needle exchange

---

et al., *Operational Issues in Syringe Exchanges: The New York City Tagging Alternative Study*, 20 J. COMM. HEALTH 111, 112–14 (1995).

[347] *See supra* notes 149–154 and accompanying text.

[348] At least one state has incorporated its concern regarding public sanction into legislation enabling a needle exchange program. The Rhode Island statute includes a statement that the act "should not be misconstrued to mean that the state of Rhode Island endorses or encourages the illegal use and/or abuse of illicit or harmful substances in any form or method of transmission whatsoever." R.I. GEN. LAWS § 23-11-18(f) (1995). Of course, this kind of official clarification is likely to have little or no effect on the perceptions of drug users.

[349] *See supra* notes 153–154 and accompanying text.

[350] As when evaluating the role of needle exchanges in reducing transmissions, we may be limited to quasi-scientific investigation—i.e., alternatives to classic experimental design—because the research in question occurs in the field rather than under laboratory control.

[351] Typically, research projects that purport to analyze needle exchange effects on both drug use rates and HIV transmission rates are far less thorough in addressing the former than in addressing the latter. Several of the studies discussed in this part serve as examples of this phenomenon.

[352] *See, e.g., infra* notes 353–354.

programs sends a signal that encourages drug injection beyond the confines of the program.

Subject to these limitations, findings generally reveal a moderate but statistically significant decline in participant injection frequency[353] or increases in injection frequency that are attributable to factors other than needle exchange participation.[354] One study of non-intravenous drug users at methadone outposts or STD clinics claimed to find "no evidence that the non-intravenous drug users . . . started intravenous drug use" upon the availability of clean needles and syringes at their sites.[355] While this finding is promising, the authors do not explain the statistical basis for the finding. Despite this vulnerability, and the limitation of the finding to nonusers in specific clinical settings, the study has been cited as refuting the concern that needle exchange programs will increase use of illegal drugs.[356]

During the five and a half year period studied by Watters et al.,[357] the median number of reported daily injections by respondents decreased significantly from 1.9 per day to 0.7 per day per person.[358] This finding appears to refute the idea that users of exchanges will increase drug use because of perceived public sanction or enablement/encouragement via provision of tools. The mean age of injection drug user respondents increased from 35.8 to 41.6 over the life of the study, and the investigators observed "a significant progressive decline in the proportion of persons who reported first injecting drugs in the previous year."[359] These observations, while not dispositive, are at least inconsis-

---

[353] *See, e.g.*, Hartgers et al., *supra* note 211, at 573 (finding that "[e]asy availability of free needles was not associated with an increase in injecting" among the participants of an interview study, and that "72% of the exchangers reported that they injected the same or less than 6 months previously compared with 49% of the non-exchangers"). *See also* Graham J. Hart et al., *Evaluation of Needle Exchange in Central London: Behaviour Change and Anti-HIV Status over One Year*, 3 AIDS 261 (1989) (reporting a statistically significant decline in frequency of injecting over the first four months following entry into a London needle exchange).

[354] *See, e.g.*, Joseph Guydish et al., *Evaluating Needle Exchange: Are There Negative Effects?*, 7 AIDS 871, 873 (1993) (noting an increase in frequency of injection attributed to an external trend that began prior to implementation of the needle exchange program studied).

[355] Van den Hoek et al., *supra* note 198, at 1356.

[356] Merrill Singer et al., *Needle Access as an AIDS Prevention Strategy for IV Drug Users: A Research Perspective*, 50 HUM. ORG. 142, 147 (1991).

[357] *See supra* note 188 and accompanying text.

[358] *Id.* at 118.

[359] *Id.*

tent with the hypothesis that needle exchange encourages new, and particularly young, drug users.[360]

Other studies that support the Watters et al. findings also suggest an alternative to the public sanction and provision of tools hypotheses. They posit instead that syringe exchanges increase drug treatment referrals or other social service referrals, which in turn can reduce the incidence of injection drug use. Syringe exchange sites have been credited with the delivery of social services beyond those directly related to drug problems.[361] These services have included tuberculosis screening, access to women's issues groups, and Bible study groups.[362] Heimer et al. also suggest that needle exchange does not promote drug use, but rather facilitates the entry of existing drug users into treatment.[363]

Although all these studies are promising, they do not yield a quality of information on which important public policy should be based. Specifically, the studies shed insufficient light on whether and how needle programs may increase illegal drug use. They do not directly examine whether the existence of government-supported needle programs may cause increased drug use outside the ranks of the needle exchange—i.e., whether publicity of the programs sets a tone of permissiveness that may encourage drug use throughout the culture. The lack of this kind of analysis is not surprising given the methodological difficulties inherent in a study of the effects of needle programs on drug use in society at large. An attempt to trace changes in drug use rates throughout a community during the life of a needle exchange is fraught with interpretive difficulties, as changes in community patterns may be associated with any of an infinite number of extrinsic variables that are difficult to control. Accordingly, the studies discussed in this part provide hints rather than compelling evidence regarding the relationship between needle programs and rates of illegal drug use.

---

[360] Heimer et al. have similarly observed no correlation between public awareness of the highly publicized New Haven needle exchange program and duration of injection drug use among needle exchange participants. Heimer et al., *supra* note 259, at 219. While this observation is far from conclusive in regard to the effect of needle programs on drug use rates, it does reveal a promising lack of change in one relevant measure.

[361] *See, e.g.*, Don C. Des Jarlais & Samuel R. Friedman, *Letters to the Editor*, 271 JAMA 1826 (1994).

[362] *Id.*

[363] Heimer et al., *supra* note 259, at 220.

Perhaps more disturbingly, one prominent research team that had found needle exchange activity to be promising in previous studies recently reported evidence suggesting the possibility that exchanges may cause some increase in drug abuse.[364] Guydish et al. interviewed participants at a San Francisco needle exchange and found that among ten percent of respondents who thought that the exchange affected the amount of drugs they used, three reported a decrease in drug use while two reported an increase in drug use.[365] While the number of participants suggesting an increase is small in this instance, and indeed is outweighed by the number suggesting a decrease in drug use, the authors observe that their findings raise concerns about the potential costs of needle exchange programs.[366] Of particular concern is a finding that "some [intravenous drug users] reported first use of injection drugs contiguous in time with first use of needle exchange," which phenomenon could be a function of a "potential role of [needle exchange programs] in recruiting new injectors."[367] The investigators suggest that further research is needed to address these concerns.[368]

More and better studies in this area will be crucial in the development of sound public policy. Until and unless needle program proponents can make convincing arguments against allegations that the programs promote and increase illegal drug use, adoption of nonprovisional, nonexperimental needle programs is ideologically suspect and politically unsound. The ideological objections are compelling at two levels: First, mitigation of future HIV transmission could be outweighed by an increase in illegal drug use, from both a public health perspective and a social welfare perspective.[369] Second, if needle programs are found to increase illegal drug use, then reductions of HIV transmissions attributable to the programs may be illusory. Specifically, declining needle sharing among existing drug users may be exceeded by increased needle sharing among new drug users—i.e.,

---

[364] Joseph Guydish et al., *Evaluation of Needle Exchange Using Street-Based Survey Methods*, 25 J. DRUG ISSUES 33 (1995).

[365] *Id.* at 38.

[366] *Id.* at 39.

[367] *Id.*

[368] *Id.*

[369] Of course, assessments of the relative values of each case of HIV transmission reduction and of additional illegal drug use are to some degree subjective. The important point here is that a needle program that encouraged illegal drug use would raise public policy questions and add credence to the arguments of those who oppose the program.

those induced to use drugs via public sanction or provision of tools.[370] This second category is particularly important because it suggests that many of the promising transmission-reduction effects noted in Part III could be rendered suspect if programs are found to encourage illegal drug use.

VI. Policy Recommendations

While research findings on needle exchange have been encouraging, they are far from conclusive.[371] We have only preliminary, imperfect knowledge, and many questions remain to be answered. The findings discussed in Part III tentatively suggest that needle programs have the potential to decrease or retard the transmission of HIV. Although promising, the research findings permit only limited inference. Methodology and research design are often seriously flawed, yielding results that are incomplete, inconclusive, and potentially misleading.

Overall, however, the findings support the continued examination of needle program efficacy. Our goal at this time should be the pursuit of research that fills the gaps in our present understanding of the likely effect of needle exchange on HIV transmission rates and illegal drug use rates. Improvement of our knowledge in these areas can be achieved by employing two basic strategies: (A) the design and adoption of provisional, researcher-friendly needle programs, the foremost mission of which is to advance our understanding of the effects of needle exchange; and (B) the development of studies that mitigate the methodological and design flaws of past research projects. Each strategy is discussed in detail below.

---

[370] *See* William N. Eskridge & Brian D. Weimer, *The Economics Epidemic in an AIDS Perspective*, 61 U. Chi. L. Rev. 733, 769 (1994) (discussing the belief of some critics that "increased availability of . . . safer mechanisms [such as needles and condoms] will only encourage the underlying activities [anal or vaginal sex and drug use], thereby canceling out the good effects" of subsidized needle exchange and condom distribution programs).

[371] *See supra* part III.

A. *The Design and Adoption of Provisional, Researcher-Friendly Needle Programs*

Policymakers should design and adopt provisional needle programs designed to be researcher-friendly in order to advance our understanding of the effects of needle exchange.[372] Given the potential value of needle programs and our need for more information regarding their precise consequences, either the unconditional acceptance or the wholesale rejection of needle exchange would be unwise.

As a practical matter, the investigation of needle exchanges requires the adoption of provisional pilot programs. These pilots are a reasonable compromise between the demands of needle exchange supporters and the objections of needle exchange detractors.[373] They permit the limited, controlled adoption of needle exchanges subject to the requirements that they be studied carefully, and they be designated clearly and expressly as test models, subject to either overhaul or abandonment should the results of preliminary studies prove their inadequacy. While some readers might consider this requirement excessively cautious,[374] under present conditions, a principal goal of providing useful, reliable information regarding program effects seems sensible.[375] The

---

[372] A provisional adoption of needle exchange for further study exists in Hawaii, where needle exchange is permitted by statute. The law requires the filing of detailed reports to an oversight committee, including data on both HIV transmission rates and drug use rates. HAW. REV. STAT. § 325-116 (1994). The law also allows termination "at any time if the program does not serve its intended purpose, presents a risk to the public health, safety, or welfare, or is no longer necessary." HAW. REV. STAT. § 325-117 (1994).

[373] Even among those who are generally optimistic about the potential of needle exchange programs, critical evaluators admit that the existing data are far from conclusive. In the wake of CDC support for needle exchanges and its call to end the ban on federal funding of the programs, the Director of HIV-AIDS Policy at CDC cautioned: "I don't think anyone is claiming that it has been proven that HIV transmission is diminished in needle-exchange-using populations. We are taking the position that the answer is—is not in yet." *Scientists Recommend Reinstatement of Needle Exchanges* (NPR Morning Edition, Mar. 9, 1995). Likewise, a spokesperson for the Drug Policy Foundation has acknowledged that the data on needle exchange effects are incomplete. Nonetheless, he notes that "[t]he ethics . . . change when we have a disease which is incurable. We cannot wait until every study is tied up with a neat knot when people are dying by our inaction." *Id.* This perspective seems eminently reasonable and supports the funding and implementation of experimental needle exchange programs.

[374] In early 1995, for example, a group of scientists urged the Clinton administration to end entirely the ban on federal funding of needle exchange programs, given the promising evidence of existing studies. Richard O'Mara, *Scientists Urge Lifting of Ban on Needle Exchanges for Addicts*, BALTIMORE SUN, Mar. 11, 1995, at 10A.

[375] Critics who believe that abstinence is the ideal means of curbing needle transmissions of AIDS may argue that even provisional, experimental needle programs should

findings cited in Part III support the implementation of needle exchanges, subject to ongoing study and monitoring.

Under this reasoning, the primary role of needle programs in the 1990s should be a research role.[376] Policymakers of all political persuasions should be willing and eager to support research programs that may ultimately vindicate or refute the logic of needle exchange. This philosophy requires that both critics and supporters of needle exchanges make concessions to the research function.[377] Critics should support the provisional adoption of exchanges, provided they can be demonstrated to function as part of a well-designed, objective study. To sustain research designs capable of providing meaningful findings, provisional adoption time frames must be long enough to generate meaningful data.[378] Likewise, supporters may need to sacrifice optimal access to needle exchange facilities in favor of

---

be rejected simply because they entail investigation of a sub-optimal policy. According to this reasoning, drug use abstinence is the most desirable means of eliminating future needle infections, therefore resources should be focused on eliminating rather than accommodating illegal drug use. From this standpoint, evidence that needle programs are likely to reduce infection rates is not compelling, because the decrease in infections attributable to needle programs is not being compared to decreases in infections attributable to either existing or potential wars on drugs.

The assumptions behind this otherwise reasonable logic are unsound. The claim that either existing or stepped-up drug prevention programs can reduce AIDS transmissions as effectively as needle programs is dubious. As Vlahov and Brookmeyer observe, "only 10% to 15% of injection drug users are in treatment at any given time, and only one half have a history of drug abuse treatment." David Vlahov & Ronald S. Brookmeyer, *Editorial: The Evaluation of Needle Exchange Programs*, 84 AM. J. PUB. HEALTH 1889 (1994) (citations omitted). This resistance to treatment programs is observed during an era in which the war on drugs has been considered a high priority. It is unrealistic to argue that further spending on prevention will have a measurable impact upon needle-sharing-induced HIV transmissions.

[376] This observation may appear callous in that it purports to place research functions over prevention functions. In response to this anticipated criticism, I respond that the prevention function, which is the ultimate goal of needle programs, cannot be served with confidence until we understand it well. As noted earlier, plausible arguments exist to suggest that needle programs could ultimately increase HIV transmission rates. *See supra* note 141. Accordingly, we cannot confidently pursue the public health goal of lowering seroconversion rates via needle exchange until we have done more to understand the precise effects of needle exchange.

[377] Ideally, these concessions are in the interests of both the critics and the supporters. Each group has posited why it believes that needle exchanges will be either harmful or beneficial. Each should expect that objectively formulated studies will ultimately prove its purportedly logical assertions to be correct. While needle exchange critics may not like the provisional implementation of programs that they see as socially destructive, they should at least be placated by the promise of information regarding the costs and benefits of needle exchange. Should critics' assertions prove correct, policy changes eliminating needle programs should justify their short-term implementation for experimental purposes.

[378] For example, the 60-day time frame of the provisional D.C. program studied by Vlahov et al. was prohibitively brief. *See supra* note 183 and accompanying text.

optimal research conditions. For example, we have observed programs in which both mandatory and voluntary HIV testing of participants was rejected because it might discourage utilization.[379] Until and unless we have convincing evidence that the programs provide net social value, needle exchange supporters should defer such access-related concerns. If HIV testing enhances the value of a study, it should be adopted.[380] In the midst of controversy among both political and academic voices, the best primary function of today's needle programs is a research function.

Given this Article's support for provisional, experimental needle exchanges, a second-tier question of financial support arises. The question of whether needle exchanges should be legalized and adopted will inevitably be tied to the issue of whether federal funding should be made available for the development of such programs. Is it enough for all levels of government to allow needle exchange experimentation pending more definitive conclusions, or is it also advisable that the federal government finance the programs?

While this question cannot be answered in a vacuum—the rational allocation of scarce resources cannot exist without the simultaneous evaluation of competing demands for funds—one crucial observation is in order here. The present ban on federal funding is overly burdensome. As noted earlier, that ban provides an escape clause, under which the funding proscription can be voided if the Surgeon General finds that needle exchanges reduce both drug abuse and HIV transmission rates.[381] This standard is oppressive. Needle exchange programs are not primarily intended to reduce drug abuse. While a needle exchange-related decline in drug abuse would be desirable, and is consistent with some needle exchange theories,[382] the primary purpose of needle exchange is the reduction of HIV transmission rates. While any increase in drug abuse engendered by the programs would cause serious policy concerns, a requirement that the Surgeon General find a decrease in abuse creates a superfluous standard that ex-

---

[379] *See supra* note 258 and accompanying text.

[380] This example presumes that sequential testing of participants for HIV over the course of a study might provide useful information. It is of course possible that researchers may believe that such testing would adversely increase attrition rates in longitudinal studies or otherwise cause more harm than good.

[381] *See supra* note 92.

[382] *See, e.g., supra* note 299–301 and accompanying text.

ceeds rational expectations of an effective needle exchange program. Accordingly, Congress should limit the effectiveness threshold to findings that the programs reduce HIV transmission rates without increasing rates of illegal drug use.

B. *Development of Studies that Mitigate the Flaws of Past Research Projects*

Investigators should develop studies that mitigate the methodological and design flaws of past research projects. In Part III, we examined research on needle exchange effectiveness in reducing HIV transmission rates, and we noted both promising results and limitations associated with the methodology and research design. In Part V, we observed that research concerning the relationship between needle exchange and illegal drug use rates has yielded only a fraction of the information needed to inform sound public policy. The value of adopting provisional, experimental needle exchange programs as recommended in the preceding subsection will depend upon the concomitant adoption by investigators of research strategies that take them beyond the limitations of past endeavors. This subsection proposes suggestions for future research projects based on the strengths and weaknesses of previous work.

The kinds of studies most promising for the future will be those that fill the knowledge gaps identified in Parts III through V. Because all quantitative and qualitative methods are subject to error, and because control is more difficult to achieve in the field than in the laboratory,[383] investigators should employ a variety of methods that are subject to different sources of error. Specifically, three categories of analysis are particularly appropriate for future research on needle exchange programs: (1) better statistical studies; (2) ethnographic studies that supplement statistical analysis; and (3) improved deductive analysis of needle exchange effects. Each of these categories is addressed individually in the remainder of this part.

---

[383] *See supra* note 172.

1. Better Statistical Studies

The effect of needle exchange on the rates of both HIV transmission and illegal drug use can be better understood by designing high-quality quantitative studies.[384] Quantitative studies of the relationship between needle exchange and HIV transmission rates should be fashioned to exploit the strengths and mitigate the weaknesses of the research examined in Part III. Among the lessons to be learned are the following:

a. *Studies should track HIV transmissions rather than proxies for HIV transmission.* Ideally, studies should track HIV transmission rates rather than imperfect proxies for HIV transmission. Studies that examine needle sharing practices or track and test needles exchanged through a needle program provide clues rather than definitive information concerning HIV transmission rates. Reductions in needle sharing do not necessarily imply concomitant reductions in the spread of HIV.[385] Testing of returned needles leaves many unanswered questions regarding the use of unreturned needles and needles from sources outside the exchange being studied. Needle return rate studies fail to provide investigators with a history of a needle's use, or to tell researchers anything about the practices associated with a needle while it remains in circulation.[386] Investigators can enhance needle return and needle tracking data by analyzing seropositivity rates of needle exchange users in comparison to seropositivity rates among nonusers, controlling for extraneous variables.[387] Seroconversion is the variable that ulti-

---

[384] *See* JONI N. GRAY ET AL., ETHICAL AND LEGAL ISSUES IN AIDS RESEARCH 165 (1995).

[385] For example, it is possible that reductions in sharing practices among persons who fail to use exclusively new needles will not be sufficient to yield commensurate reductions in HIV transmission if even this reduced incidence of sharing is sufficient to evoke a transmission among individual users. It would be particularly important to test this hypothesis if investigators were to observe widespread reductions in sharing practices but few instances in which needle sharing is completely eliminated.

[386] This problem is exacerbated as the time span for returns being studied increases. The longer a needle is in potential circulation, the greater the opportunities for a wide variety of uses. The problem, of course, is that exchange users cannot be expected to return needles soon after their original distribution.

[387] For discussion of the failure of prior studies, including some of the studies discussed in part III, to invoke adequate control, see Noreen V. Harris et al., *Assessing the Efficacy of Needle Exchange Programs: An Epidemiological Perspective*, in PROCEEDINGS: WORKSHOP ON NEEDLE EXCHANGE AND BLEACH DISTRIBUTION PROGRAMS, Panel on Needle Exchange and Bleach Distribution Programs, Sponsored by Commission on Behavioral and Social Sciences and Education, National Research Council, and Institute of Medicine 155, 159–60 (1994).

mately matters, and other measures used in isolation leave too many questions unanswered to inform important public policy decisions.[388]

b. *Studies should focus on long-term effects of needle exchange.* Studies should be designed to examine both long-term and short-term effects of needle exchange on seropositivity rates. Most of the studies examined in Part III examined behaviors or other proxies for seroconversion over extremely short periods of time. We need to understand the long-range implications of needle exchange to determine whether the programs are the best use of scarce resources. For example, long-term studies could reveal that needle exchange programs tend to delay rather than avert seroconversion. This might be true if, for example, participants only occasionally utilized needle exchange programs. It is plausible that occasional use of needle exchange might delay the transmission of HIV without reducing the number of persons who ultimately contract the virus.

Alternatively, long-term studies may unveil the less dramatic but perhaps more likely scenario that long-term reductions in seroconversion are small in comparison to short-term reductions, as some occasional needle exchange users experience a delay rather than an avoidance of infection. Were any of these examples to be demonstrated by analysis of data over a long period, policy makers might reasonably decide to invest in alternative AIDS prevention programs found to be more effective in saving lives.[389]

c. *Studies should be designed so that causality can be inferred and comparisons can be made.* Studies should be designed to provide findings that can support inferences regarding a relationship or causality between needle exchange programs and hypothesized outcomes. We observed in Part III that many studies are designed so that their utility to lawmakers is limited. For example, research by Johnson et al. in Dublin revealed short-term needle sharing rates within one

---

[388] Seroconversion rates are the ultimate dependent variable of concern, whereas tracking information and return rates are only secondary proxies that are reasonably presumed to have some causal connection to seroconversion rates. Seroconversion rates should, therefore, be considered primary data superior to more attenuated, secondary data. Nonetheless, secondary data on needle exchange rates, needle contamination, and needle history will remain important components in fleshing out the effect of needle exchange programs on slowing the transmission of HIV.

[389] These hypotheticals are only meant to serve as examples of the ways in which long-term studies may provide findings that lead policy makers in a direction different than the direction deduced from short-term data.

group of exchange users.[390] Such information is of little use in the formation of public policy for two reasons: it yields no comparative insight and it provides no opportunity for the inference of causality.

Research that is most helpful to lawmakers provides relative, rather than absolute, information. To this end, needle exchange studies should entail some form of comparison. Needle sharing rates among exchange users become more meaningful when they are compared, for example, to needle sharing rates of exchange users prior to participation in the exchange, or to a group not participating in needle exchange, the composition of which has been controlled to ensure against the effects of extraneous variables.[391]

Ideally, studies should also be designed to permit the confident inference of cause and effect. The classic approach to this end is scientific experimentation, employing random sampling,[392] random assignment into treatment and control groups,[393] and statistical analysis of treatment effects.[394] As lawmakers adopt provisional, experimental needle exchanges designed primarily for study as recommended in subsection (1), they should authorize and encourage the utilization of experimental designs that permit, within strict ethical standards of human behavioral research, inference of causal relationships.

d. *Studies should examine effects of needle exchange on illegal drug use.* Investigators should devote far more effort to higher quality empirical study of the relationship between needle programs and drug use rates. To date, the only encouraging findings that are even moderately persuasive indicate that needle programs may not increase illegal drug use among program participants.[395] Further analysis is needed to help determine whether needle programs increase illegal drug use in the community at large, and not just among program participants.

Studies that simply record changes in community-wide illegal drug use rates over the life of a needle exchange program are

---

[390] *See supra* text accompanying notes 239–244.

[391] For discussion of research challenges of this variety, see Charles S. Reichardt, *The Statistical Analysis of Data from Nonequivalent Group Designs*, *in* THOMAS D. COOK & DONALD T. CAMPBELL, QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD SETTINGS 147–205 (1979).

[392] COOK & CAMPBELL, *id.* at 75.

[393] Neale & Liebert, *supra* note 169, at 53–55.

[394] *Id.* at 13–34.

[395] *See supra* notes 353–360 and accompanying text.

also of limited value.[396] Better research projects would compare changes in drug abuse rates in needle exchange communities with changes in comparable non-exchange communities,[397] and would use analysis of variance techniques to provide clues regarding the degree to which each of a number of posited variables (including the presence of needle exchange programs) account for an observed variation.

e. *Studies should be repeated in different contexts to ensure universal applicability.* Studies noting success or failure in one situation should be replicated in other contexts, especially when situational dynamics appear likely to diverge. As noted in Part III, the findings of any needle exchange study may be context-specific.[398] Variables across different populations may affect the utility and characteristics of optimal programs.[399] For example, legal and cultural environments are likely to affect the feasibility of needle programs. Findings from the Netherlands may not be applicable to policy in the United States. These two nations[400] have adopted fundamentally divergent approaches to drug use—a "harm reduction" approach and a "legal confrontation" approach, respectively.[401] The United States' legal confrontation and treatment policy criminalizes the use of certain drugs and employs police enforcement to effect reductions in drug use.[402] The harm reduction approach in the Netherlands combines decriminalization with

---

[396] Such studies identify trends without permitting either an inference of causality or an illuminating comparison.

[397] The use of comparable non-exchange communities can help control for the effects of unforeseen extraneous variables. It is possible, for example, that drug abuse rates will grow throughout a region during a period that happens to coincide with one locality's adoption of a needle exchange program. This increase in drug abuse might be attributable to any number of political, social, and economic phenomena.

Data indicating that needle exchange communities consistently experience greater increases in the incidence of drug abuse would, however, lead to the conclusion that these programs contribute to drug abuse. If the results are comparable to other non-exchange communities, the changes are likely attributable to sources other than needle exchange activity.

[398] *See supra* note 169 and accompanying text.

[399] *See supra* note 170 and accompanying text.

[400] The United States and the Netherlands are, of course, not the only countries that have adopted either of the two generic approaches discussed here. While policy in most countries resembles the U.S. policy, Switzerland has addressed the problem using the Dutch harm reduction approach. For a discussion of Swiss policy and Zurich's needle park, see Mike Huggins, *Europe's Heroin Haven*, INDEPENDENT (London), Feb. 16, 1995, at 22.

[401] Pascal Imperato, *Editorial, Needle Exchange and the Prevention of AIDS*, 14 J. COMMUNITY HEALTH 61, 62 (1989). For further discussion of these approaches, see *supra* notes 110–111 and accompanying text.

[402] *See* Imperato, *supra* note 401.

public policies aimed at reducing the pernicious effects of drug abuse rather than the incidence of drug abuse itself.[403]

These differences in the legal and cultural environments of the United States and the Netherlands are likely to affect the viability and success of needle exchanges. For example, the threat of criminal enforcement in the United States could increase exchange patron wariness. If, as studies suggest, regular and exclusive resort to needle exchange for one's supply of needles is a necessary element of an effective program, distrust engendered by a culture of legal confrontation is likely to undermine the effectiveness of exchange programs.[404] The absence of such a threat in the Netherlands may render needle programs more viable in that country, so that success rates in Dutch studies may not translate into accurate predictions of success in the United States.[405]

Accordingly, results of studies within a particular context should be scrutinized carefully to determine their external validity.[406] When doubts about external validity arise, studies should be replicated across divergent settings.

2. Ethnographic Studies that Supplement Statistical Analysis

Ethnographic studies can supplement the statistical analysis recommended in subsection (1) with more intimately observed narrative information that helps explain how individuals respond to needle exchange programs. Investigators should consider less traditional, nonscientific methods of obtaining information to verify and supplement scientific findings. Specifically, they should fortify quantitative approaches with qualitative, ethnographic methods that can help to explain behavioral patterns and trends.

Ethnography is a qualitative research methodology[407] used by historians and anthropologists to understand events within their

---

[403] *See id.*

[404] It is natural that a police presence or the belief that a police presence is possible would have a chilling effect on needle exhange programs in a culture of legal confrontation.

[405] Notwithstanding this reasonable hypothesis, the needle programs studied to date in the United States and the Netherlands both show considerable promise. *See supra* part III.

[406] For discussion of external validity problems, see *supra* notes 169–170 and accompanying text.

[407] Susan C.M. Scrimshaw, *Bringing the Period Down: Government and Squatter Settlement Confront Induced Abortion in Ecuador, in* MICRO AND MACRO LEVELS OF

cultural settings[408] and to reconstruct "the distinctive mentalities" of the relevant actors.[409] Anthropological ethnographers achieve these ends by cultivating close personal relationships with informants.[410] By observing, speaking with, and living among informants in their own cultural settings,[411] the ethnographer aspires to combine an insider's perspective of the culture's dynamics with illumination derived from anthropological methodology.[412] She or he constructs a narrative capable of explicating social interactions, behaviors, and cultural patterns.[413]

The advantages of the ethnographic process in supplementing quantitative analyses are, ironically, a function of the methodology's lack of classic scientific rigor.[414] One scholar observes that "a purely objective ethnography is impossible, because ethnography is determined by context, rhetoric,[and] the institutions in which it is written . . . ."[415] Postmodern ethnography is accordingly described as "poststructural, processual, interpretive, symbolic, cultural, and hermeneutic."[416] It is unabashedly subjective,

---

ANALYSIS IN ANTHROPOLOGY: ISSUES IN THEORY AND RESEARCH 121, 123 (Billie R. DeWalt & Pertti J. Pelto eds., 1985).

[408] Anita F. Gelburd, The Harvard Report of 1945: An Historical Ethnography 4 (1994) (unpublished Ph.D. dissertation, University of Pennsylvania) (on file with the *Harvard Journal on Legislation*).

[409] RHYS ISAAC, THE TRANSFORMATION OF VIRGINIA, 1740–1790, at 323 (1982).

[410] JAMES P. SPRADLEY & DAVID W. MCCURDY, THE CULTURAL EXPERIENCE: ETHNOGRAPHY IN COMPLEX SOCIETY 41 (1972).

[411] *Id.* at 45.

[412] This methodology is typically referred to as "participation-observation fieldwork." *See, e.g.*, Dwight Conquergood, *Rethinking Ethnography: Towards a Critical Cultural Politics*, 58 COMM. MONOGRAPHS 179, 180 (1991).

[413] For an explanation of the specific narrative approaches that are used by ethnographers seeking to construct rigorous, accurate descriptions, see Robert Aunger, *On Ethnography: Storytelling or Science?*, 36 CURRENT ANTHROPOLOGY 97, 104–06 (1995).

[414] For example, whereas scientists employ double-blind testing to achieve objectivity of both observer and subject, ethnography has been described as subject to both "anthropologists' preconceptions and those of their subjects of study." Paul A. Roth, *Ethnography without Tears*, 30 CURRENT ANTHROPOLOGY 555, 555 (1989) (citation omitted).

Another source of subjectivity is embodied in the writing of a field study. As one commentator observes, "the fieldworker must . . . put into words what was learned of a culture so that a representation of sorts may result . . . . Yet any claim to directly link fieldwork . . . to the ethnography itself, unmediated or untransformed by narrative conventions, will not hold." JOHN VAN MANNEN, TALES OF THE FIELD: ON WRITING ETHNOGRAPHY 7 (1988). This limitation results from the notion that "culture or a cultural practice is as much created by the writing (i.e., it is intangible and can only be put into words) as it determines the writing itself." VAN MANNEN at 6.

[415] Kevin K. Birth, *Reading and the Righting of Writing Ethnographies*, 17 AM. ETHNOLOGIST 549, 550 (1990).

[416] Norman K. Denzin, *Review Symposium on Field Methods*, 18 J. CONTEMP. ETHNOGRAPHY 89, 90 (1989) (citation omitted).

and many postmodern ethnographers acknowledge and extol the methodology's insider perspective.[417]

While the disadvantage of immersing an observer into the context being studied is the lack of objectivity engendered by intimacy,[418] the countervailing advantage is a capacity for close observation, a function sacrificed by traditional research that aspires to an ideal of perfect scientific impartiality.[419] Ethnography can reclaim an important element forfeited by the scientists whose studies have dominated the issue of needle exchange. It can achieve perceptions and insights likely to appear only from close observation within a culture.

The suggestions made by investigators in discussing the implications of their traditional findings, subject as they are to methodological limitations, can be supported or rebutted by case forms[420] of ethnographic analysis,[421] which tell more closely the stories of individuals, their acts, and their motives.[422] Case analyses may also serve to generate better informed hypotheses, thereby improving the research questions that scientific and quasi-scientific

---

[417] For a concise sense of the subjectivity of ethnography, see RENATO ROSALDO, CULTURE AND TRUTH: THE REMAKING OF SOCIAL ANALYSIS 37 (1989). For a discussion of the ways in which ethnography can arguably remain objective, see Aunger, *supra* note 413.

[418] While lack of objectivity is certainly a risk of ethnographic research, it is also a risk in ostensibly scientific research, as even the best-intentioned scientific investigator is susceptible to selective observation and selective interpretation. For discussion of objectivity problems attendant upon all forms of research and of the mechanisms for reducing bias, see SPRADLEY & MCCURDY, *supra* note 410, at 13–21.

[419] Of course, the rigors of science are crucial to the execution of good research. As noted in the previous subsection, projects that are intended to be classically scientific should aspire to more, rather than less, rigor. The comments in this subsection are meant to apply to entirely different types of analysis which are unabashedly nonscientific. Although science in its purest form is more objective, ethnography nonetheless can enhance insight and understanding. Perception may be sacrificed when investigators who strive for objectivity attain a desired distance from the subjects they are studying, only to lose the advantages that may be derived from close scrutiny. Science and ethnography can be complementary—each providing the advantages that the other must sacrifice by its nature.

[420] The reference to case analysis here is a broad one. It includes both the tracking of individuals to develop typical stories and the statistical study of aggregate numbers of cases over extended periods of time. While statistical case analysis is more anonymous and objective than the more personal, subjective approaches taken in the qualitative analysis of cases, it represents an important departure from the studies discussed in part III.

[421] *See* Aunger, *supra* note 413, at 97 ("The interpretive or narrative approach . . . aggregates over events to trace the . . . development of a single case."). A rough analogy can be made to the documentary form in film-making, from which tremendous insight can be derived—insight occupying a different dimension than the insight derived from statistical analysis.

[422] For discussion of the use of story as it has evolved in historical research, see David Samuels, *The Call of Stories*, LINGUA FRANCA, May-June 1995, at 35.

investigators choose to investigate.[423] Long-term tracking of participants' lives, or relevant aspects of their lives, adds a new and important dimension to the literature; it enhances our understanding of human experience through the highly concentrated, focused observation of individuals. Employing anthropological ethnography, investigators can analyze intimately "sets of actions/relations, roles/statuses, and patterns/structures," which "pertain to the way in which individuals or groups of individuals live their lives."[424]

3. Improved Deductive Analysis of the Effects of Needle Exchange Programs on Attitudes and Behaviors

In examining the likely ramifications of needle programs in terms of illegal drug use, we should not summarily dispose of rational discourse as a means of predicting likely cause and effect. Particularly when empiricism is limited by methodological difficulties and unavoidable potential for error, the ability of scholars and commentators to predict effects through the use of reason is an important supplementary tool.

Unfortunately, in the largely political debate over needle exchange programs, careful and considered analysis of what might be the logical effects of needle programs has been extremely rare. Careful reasoning regarding the effects of needle exchange on drug abuse rates can begin by understanding the nature of the two crucial arguments—public sanction and provision of tools—against such programs.[425] Discourse regarding the public sanction theory should focus on the degree to which the legal climate influences the decisions to use drugs. Researchers should ask whether drug abuse is somehow different from other undesirable behaviors in its susceptibility to threat of punishment. One ob-

---

[423] As investigators observe promising statistical trends in HIV patterns among injection drug users participating in needle programs, they are challenged to break success stories into elements of success. For example, they might try to understand the factors that appear to differentiate successful participants (i.e., those who maintain seronegativity) from unsuccessful participants (i.e., those who become seropositive during their participation). While quantitative analysis can be used to test any number of hypotheses in regard to this question, the creation of intelligent hypotheses can benefit from qualitative analysis—i.e., ethnographic case study. Hypothesis generation is yet another role that ethnography can and should play in future needle exchange research.

[424] GOPALA SARANA, THE METHODOLOGY OF ANTHROPOLOGICAL COMPARISONS: AN ANALYSIS OF COMPARATIVE METHODS IN SOCIAL AND CULTURAL ANTHROPOLOGY 94 (1975).

[425] *See supra* text accompanying notes 151–156.

vious distinction between continuing drug use and the decision to try using drugs for the first time is the compulsion of addiction among many ongoing users. Addiction would logically be expected to be more resistant to legal and other extrinsic threats.

A second critical area of analysis regarding the public sanction argument is whether the government's provision or support of needle exchange facilities is likely to be viewed as sanctioning drug use. Those who understand the function of needle exchange may disassociate such programs from the question of legality of drug use, considering the two to be separate and distinct issues. The rationality of such a division would be consistent with the extent of our ability to hold two beliefs simultaneously—that drug use should be criminalized and that needle exchange programs should be adopted to thwart the spread of HIV because drug users will continue to break the law. The recognition of the former fact does not necessarily suggest that illicit drug use is or should be sanctioned. On the other side, we must consider whether drug users or potential drug users are likely to deduce that support for needle exchange implies a sanction of illicit drug use. While needle exchanges do not necessarily sanction drug use, they could conceivably imply such sanction, and the existence of such exchanges might alter some people's conceptions of the law's severity in regard to drug abuse.

Likewise, policymakers should engage in careful and considered discussion of the rationality of the provision of tools theory. They should examine whether lack of lawful access to needles and syringes reduces intravenous drug use. Reducing such lawful access would seem more likely to impede initial use of drugs than future use by experienced injectors, as the latter group is more likely to have old paraphernalia or network access to new but illegal paraphernalia. Nonetheless, it is an area that requires study.

Participants in the debate must also be careful to avoid suspect metaphors and analogies.[426] Reducing the debate to groundless assertions is not in the interest of sound policy. Just as critics of

---

[426] One director of public health has been quoted, for example, as suggesting that needle exchanges do not encourage illegal drug use, reasoning that "providing someone with a fork doesn't make him eat more." Ernst C. Buning et al., *Amsterdam's Drug Policy and its Implications for Controlling Needle Sharing*, 80 NAT'L INST. ON DRUG ABUSE RES. MONOGRAPH SERIES 59, 72 (1988). While this statement is undoubtedly true, it is not entirely relevant to the provision of tools theory. While it is hard to imagine a person starving for lack of a fork, it is easy to imagine a person denying a desire to indulge in injection drug use for lack of a needle.

needle programs should provide persuasive arguments in support of their fears, so supporters should construct careful, convincing arguments to dispel those fears.

The questions posited in this subsection are intended to be exemplary rather than exhaustive. Aspects of these questions can and should also be addressed using fairly traditional research methodology. The point here is not that such methodologies are useless; it is rather that these methodologies are imperfect, and that other ways of knowing and developing insights can be used to complement more traditional research findings. We should not dismiss the value of meticulous rational discourse in the process of illuminating public policy questions. The key is to ensure a thorough and meticulous engagement in addressing these issues. The merits of rigorous debate and thoughtful reasoning should not be undervalued.

VII. CONCLUSION

The decision to develop and implement a needle exchange program should be based on facts rather than untested assumptions. Sound public policy decisions that should be predicated upon reliable information are made instead in reaction to political posturing which capitalizes on ungrounded fears, simplistic moralistic conclusions, or ignorance. Proposed needle exchanges are susceptible to either acritical support or knee-jerk disapproval, in part because emotions run so high among both supporters and opponents.[427] This Article has been an attempt to persuade policymakers, lawyers, legal scholars, and citizens that the viability and wisdom of needle exchange programs can and should be assessed on the basis of careful, methodologically sound study.

Toward achieving this end, I have recommended that lawmakers provide temporary statutory approval of experimental needle exchange programs with the primary purpose of studying their effects. Emphasis is placed here on the provisional nature of this approval, as well as its experimental function. We must fill the gaps in our present understanding of the social effects of needle exchanges before we decide to accept or reject them; the only rational approach to achieving this end is the cautious encour-

[427] For the contentions on both sides, *see supra* part II.

agement of high-quality investigation. While we need needle exchange programs from which to learn, their function should be considered predominantly investigative until and unless their benefits are more conclusively validated. Likewise, because many of their effects remain to be seen, statutory approval of experimental needle exchange activity should be provisional. If their utility is vindicated, then and only then should needle exchange programs receive permanent authorization.

The limitations in our present knowledge have been discussed in great detail in the preceding parts. I have identified and analyzed three vital areas of inquiry: whether needle exchanges appear to reduce or delay total HIV transmission via shared needle use;[428] the nature and characteristics of those exchanges that seem to be most promising;[429] and whether needle exchanges appear to increase drug use, either among program participants or within the community at large.[430] The data that exist to date invite a cautious optimism regarding the value of needle exchange programs. While the literature suggests that needle exchange programs can prevent some future HIV infections, the quality of research design varies widely from study to study. Findings are based on quantitative rather than qualitative data. They rely on statistical analysis of information that has been taken from the field and is therefore difficult to control. The studies generally depend on proxies for HIV transmission rates rather than actual measures of HIV transmission itself; inference of causality is frequently speculative. Within this maze of potential hazards, the care used by investigators to maximize reasonable opportunities for legitimate inference varies substantially.

Investigators should exercise care in choosing research methodology and design. Furthermore, qualitative research that examines the progression of individual case studies will provide valuable narrative information that may help to explain and elucidate the ways in which needle exchanges motivate various kinds of behavioral changes. Because the logic of needle exchange as a means of reducing HIV transmissions is based on behavioral assumptions, this kind of qualitative information is extremely valuable. In light of the scientist's disposition to favor the quantitative over the qualitative, special efforts must be made

---

[428] *See supra* part III.
[429] *See supra* part IV.
[430] *See supra* part V.

to ensure that the latter mode of research is given an appropriate emphasis.

These observations apply not only to research that seeks to determine whether needle programs can reduce future HIV transmissions, but also to research that attempts to understand the qualities and characteristics of optimal needle exchange programs. Moreover, research that addresses any differentiation of success factors under varying sets of conditions or assumptions will provide an important basis for public policy development in areas where needle exchanges have proven themselves worthy of extended or permanent adoption. The recommendations in Parts III and IV suggest a two-tier process in the scrutiny of needle exchange effectiveness: high quality studies using a variety of legitimate and complementary techniques, in order to assess (1) the value, if any, of the needle exchange concept, and (2) the ways to optimize this value. While many research projects of the second variety are likely to follow and depend upon positive findings from research projects of the first variety, studies can be designed in which both kinds of data are examined simultaneously. As we have seen in the preceding discussions, some of the best research to date has yielded both validation of, and design recommendations for, needle programs.

Our weakest area of understanding in regard to needle exchange programs remains our knowledge about the effects these programs have on illegal drug use. As we observed in Part V, the absence of persuasive data in this area is not surprising. Many of the studies that do exist are limited by their confinement to an examination of program effects on participants, leaving open the question of effect on drug usage outside the confines of the program. This remains a topic that will be of critical importance in the future public policy debate.

Part VI recommends three priorities that focus on filling the gaps in our knowledge of needle exchange programs. First, we need greater numbers of quantitative research projects that utilize sophisticated statistical tools to try to understand complex field phenomena that are likely to reflect multivariate causes. Researchers and grant providers should especially pay increased attention to the methodologically improved study of needle program effects on illegal drug use, so that the quality of these projects begins to match the quality of studies that have already been done concerning needle program effect on HIV transmission rates. While proponents of needle programs may see these

studies as a low priority, preferring instead to devote scarce resources to projects that improve delivery of the programs that exist, a failure to examine carefully the effects on illegal drug use would be closed-minded and shortsighted. As long as plausible public sanction and provision of tools arguments exist, needle programs will remain both ideologically suspect and politically controversial. Second, investigators should attempt to mitigate the imperfections of empirical field research by utilizing complementary research techniques that can help to verify or disprove quantitative findings that must be considered tentative by virtue of any inherent methodological shortcomings. Part VI briefly suggests some less quantitative methods that might qualify under this heading. Finally, commentators who are not utilizing scientific or social scientific methods, either traditional or innovative, should back their policy assertions by carefully formulated deductive reasoning. Much commentary on needle exchange, particularly in regard to predicted effects on illegal drug use, is posited in the form of unsupported and highly speculative conclusions. While scientific and quasi-scientific methods are not the only avenues toward knowledge, those using alternative approaches must nonetheless be systematic and meticulous in the reasoning of their arguments.

All of these observations regarding important areas of needle program study are specific recommendations that share a basic common philosophy—that this public policy issue should be resolved based on the highest quality of knowledge and understanding that we can attain, and not on ignorant speculation or any political concessions made in deference to such ignorance. This Article has been an effort to catalog our understanding of needle exchange to date, as well as the holes in our understanding. It is intended to provide a road map, comprised of recommendations for the effective development of knowledge bases upon which intelligent policies can be based. By authorizing provisional, experimental programs for rigorous scrutiny, lawmakers build public policy on rational discourse rather than political expediency.

ARTICLE

INTERSTATE BANKING AND BRANCHING UNDER THE RIEGLE-NEAL ACT OF 1994

MARK D. ROLLINGER*

In this Article, Mark Rollinger examines the Riegle-Neal Act, legislation that will allow banks to branch interstate, at least in states that do not opt out of its provisions. In surveying the history of banking regulation, including past efforts of banks to circumvent geographic restrictions, the Article shows that the Riegle-Neal Act was less revolutionary than commonly perceived. The Article explores the issues and controversies surrounding the passage of the legislation, including its potential effects on industry competition, access to credit, and bank safety and soundness. The Article then analyzes how these same issues will influence the debate at the state level concerning whether to opt out. Further, the Article focuses on the potential effect this legislation will have on the banking industry, concluding that it may lead to large cost savings but will not necessarily lead to the demise of community banks or bank holding companies. Ultimately, the Article finds that the legislation was a skillful compromise, building on the present structure while maintaining a large local influence on banking regulation.

Two central concerns in the history of American banking regulation are what banks do and where they do it. Both have been fiercely debated by successive generations of bankers and regulators, and the legal doctrine surrounding them has developed in stages from extreme restrictiveness to today's relative permissiveness. The regulatory structure in this area continues to evolve. In 1994, Congress passed landmark legislation accelerating the movement toward banking and branching across state lines,[1] and has recently turned its attention to bills introduced in both houses[2] that would lift the limitations on bank securities powers imposed by the Glass-Steagall Act of 1933.[3] Although it is too early to assess this latest round of bank powers discussions, the time is

* Associate, Cleary, Gottlieb, Steen & Hamilton, New York, New York (not admitted to practice). B.A., Yale College, 1991; J.D., Harvard Law School, 1995.

[1] Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994, Pub. L. No. 103-328, 108 Stat. 2338 (1994) (codified in scattered sections of 12 U.S.C. (1994)) [hereinafter Riegle-Neal Act].

[2] Financial Services Competitive Act of 1995, H.R. 18, 104th Cong., 1st Sess. (1995); Depository Institution Affiliation Act of 1995, S. 337, 104th Cong., 1st Sess. (1995).

[3] The Banking (Glass-Steagall) Act of 1933, ch. 89, 48 Stat. 162 (codified in scattered sections of 12 U.S.C.), separates commercial banking from investment banking. Section 16 prohibits national banks from underwriting, selling, or dealing in securities. 12 U.S.C. § 24 (Seventh) (1994). Section 20 forbids affiliations between banks and securities firms. 12 U.S.C. § 377 (1994).

ripe for an analysis of the recent debate over geographic restrictions, which began in 1993 and culminated in the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 (the Riegle-Neal Act or the Act).

A review of the terms of this debate and its resolution at the federal level—as well as a look at how the arguments on both sides are now being redeployed at the state level as states decide how to respond to the Act's opt-out provision[4]—suggests the persistence of uncertainty over how to draw geographic lines around banking operations in the United States. For two hundred years, the questions of precisely what relationship a bank bears to its community and how far that community extends have remained unanswered. The struggle to define an appropriate banking "area" and the enduring belief that this area ought to be restricted in some way (for example, to a county or state or group of contiguous states) are unique to the United States. No other industrialized nation places geographic restrictions on its banks.[5] Nor has the United States sought to confine geographically other types of enterprises, including those that compete with banks, or burden them with comparable responsibilities to their local communities. Although the Riegle-Neal Act is a significant step toward eliminating unnecessary constraints and modernizing the regulatory structure, it does not mark a sharp break with the past. Banks remain the object of considerable local concern, and the states retain much of their regulatory authority. Thus, rather than definitively resolving the tension between the "local" and "global" banking models, the Riegle-Neal Act simply represents a new political compromise in the continuing effort to balance the need for local control over banks with market forces that favor enlarging banks' geographic spheres.

Broadly speaking, the Riegle-Neal Act effects two changes. First, it provides that as of June 1, 1995, a bank holding company (BHC)[6] may acquire a bank in any state, even a distant one,

---

[4] Section 102 of the Riegle-Neal Act provides that, before the cut-off date of June 1, 1997, a state may enact a law that effectively exempts banks in that state from the interstate branching provisions of the Act. *See infra* note 356 and accompanying text.

[5] For example, banks may operate nationwide in Germany, France, Britain, Canada, and Japan. *See infra* notes 319–327 and accompanying text.

[6] A bank holding company is defined as "any company which has control over any bank." 12 U.S.C. § 1841(a)(1) (1994). Control is defined as owning, controlling, or having the power to vote 25% or more of any class of the bank's voting securities; controlling the election of a majority of the bank's directors or trustees; or exercising a "controlling influence over the management or policies of the bank." 12 U.S.C. § 1841(a)(2) (1994).

irrespective of that state's laws concerning interstate acquisitions. This provision repeals the Douglas Amendment[7] to the Bank Holding Company Act of 1956 (BHCA), one of the pillars of American banking law. By prohibiting a BHC from acquiring a bank in another state unless that state explicitly authorized bank acquisitions by out-of-state BHCs,[8] the Douglas Amendment effectively transferred to the states all authority to determine the extent of BHC expansion across state lines. Maine was the first state, in 1975, to allow entry by out-of-state BHCs.[9] By 1994, every state except Hawaii had authorized acquisition of in-state banks by out-of-state concerns.[10] Thus, the provision of the Riegle-Neal Act that replaces the Douglas Amendment with a blanket federal rule authorizing BHCs to stretch as broadly across the nation as they wish, whether states permit entry or not, actually changes very little. It simply confirms—and codifies at the federal level—the free-entry rule already adopted by the vast majority of states. At most, the interstate banking provision, which does not allow for opt-out on the part of the states,[11] only removes from the states the option of insulating their banks from the possibility of acquisition by an out-of-state BHC—an option that only one state in the Union is currently exercising.

Although the second major provision of the Act appears to mark a significant change in banking regulation, it is also less innovative than it appears. This provision concerns interstate

---

[7] 12 U.S.C. § 1842(d) (1994).

[8]

> [N]o application . . . shall be approved under this section which will permit any bank holding company or any subsidiary thereof to acquire, directly or indirectly, any voting shares of, interest in, or all or substantially all of the assets of any additional bank located outside of the State in which the operations of such bank holding company's banking subsidiaries were principally conducted on July 1, 1966, or the date on which such company became a bank holding company, whichever is later, unless the acquisition of such shares or assets of a State bank by an out-of-State bank holding company is specifically authorized by the statute laws of the State in which such bank is located, by language to that effect and not merely by implication.

12 U.S.C. § 1842(d) (1994).

[9] Arthur E. Wilmarth, *Too Big to Fail, Too Few to Serve? The Potential Risks of Nationwide Banks*, 77 IOWA L. REV. 957, 977 (1992).

[10] Of the 49 states that permit acquisition by out-of-state BHCs, 34 allow entry by BHCs from any state in the nation, while the remaining 15 allow entry only by BHCs from certain states, usually those in the same geographic region. GENERAL ACCOUNTING OFFICE, INTERSTATE BANKING: BENEFITS AND RISKS OF REMOVING REGULATORY RESTRICTIONS, GAO/GGD-94-26, Nov. 2, 1993, at 23 [hereinafter GAO 1993]. *See infra* note 53 and accompanying text.

[11] Indeed, an opt-out clause would eviscerate the interstate banking provision of the Riegle-Neal Act. A state that opted out would be governed by the Douglas Amendment.

branching (as opposed to banking).[12] Effective June 1, 1997, the Riegle-Neal Act will repeal another fundamental piece of American banking legislation, the McFadden Act of 1927.[13] In enacting the McFadden Act, Congress authorized the states to determine whether, and to what extent, banks should be allowed to branch. Although strictly speaking the McFadden Act only applied state branching restrictions to national banks,[14] it was interpreted to mean that national banks could branch at most statewide, and in no case across state lines.[15] In contrast, the Riegle-Neal Act will allow BHCs owning banks in more than one state to merge those banks into a single interstate branch network.[16] In other words, a banking organization wishing to operate in multiple states will no longer be required to utilize the BHC structure, whereby it must maintain separately incorporated banks in its different states of operation, all of them unified under the umbrella of a common BHC. Instead, for the first time in American history, it will be possible for a single national bank headquartered in one state, having only one charter, to open branches in other states, whether nearby or distant. In this respect, the Riegle-Neal Act is revolutionary. Interstate branching represents a bigger step in the direction of the "global" banking model than does mere interstate expansion via a BHC structure. On the other hand, in view of the two-and-a-half year waiting period before the branching provision takes effect, as well as the opt-out clause allowing states to prohibit interstate branching if they choose, the branching component of the Riegle-Neal Act may be less radical than it

---

[12] A branch is defined as "any branch bank, branch office, branch agency, additional office, or any branch place of business . . . at which deposits are received, or checks paid, or money lent." 12 U.S.C. § 36(j) (1994). Unlike an independent bank subsidiary, a branch is not separately incorporated; it does not have its own board of directors, by-laws, capital requirements, etc. A branch is simply an arm of the main bank.

[13] 12 U.S.C. § 36 (1994).

[14]

> A national banking association may, with the approval of the Comptroller of the Currency, establish and operate new branches: (1) Within the limits of the city, town or village in which said association is situated, if such establishment and operation are at the time expressly authorized to State banks by the law of the State in question; and (2) at any point within the State in which said association is situated, if such establishment and operation are at the time authorized to State banks by the statute law of the State in question by language specifically granting such authority affirmatively and not merely by implication or recognition, and subject to the restrictions as to location imposed by the law of the State on State banks.

12 U.S.C. § 36(b) (1994).

[15] EDWARD L. SYMONS, JR., & JAMES J. WHITE, BANKING LAW: TEACHING MATERIALS 98 (1984).

[16] Riegle-Neal Act § 102.

appears. In effect, it does little more than remand to the states the question of interstate branching.

This Article will examine the Riegle-Neal Act in detail, discussing the origins and functions of many of its provisions, making some predictions as to its probable consequences, and, most importantly, showing how the new legislation fits into the long (and uniquely American) struggle over how to place geographic contraints on banks. Thus, Part I of this Article will explore the history of bank expansion and geographic limitations. Through a survey of the methods that banks have used to circumvent such limitations, Part II will demonstrate that, to a considerable extent, the United States already had a system of interstate banking when Congress began to reconsider the issue in 1993–94. Part III will review the terms of the most recent debate over interstate banking and branching and will critically assess the arguments on both sides. Part IV will analyze the outcome of this debate, the Riegle-Neal Act, and will demonstrate that the new legislation represents a victory not only for the free market but also for states' rights, in that it grants considerable discretion to the states to determine how wide a geographic field banks should be permitted to occupy. Finally, Part V will explore the ramifications of the new rules and offer some thoughts about the future. The overall objective of this Article is to clarify the relationship between geography and banking—and to explain why American regulatory authorities, unlike their counterparts abroad, have insisted so strenuously over the decades that one cannot be considered independent of the other.

I. The History of Bank Expansion and Geographic Restrictions

A. *The Nineteenth Century: The Unit Banking System*

Distrust of powerful financial institutions has been part of American folklore since the founding of the Republic. It has also played a significant role in American political history. In the early nineteenth century, Thomas Jefferson and James Madison argued against concentrated banking power.[17] Consistent with their view, the early American banks were state-chartered banks

[17] Wilmarth, *supra* note 9, at 970.

scattered across the country; these multiplied rapidly during the first decades of the nation's history.[18] Only the central bank, the Bank of the United States, had a national charter and a nationwide branching system.[19]

One of the main arguments advanced in Congress against renewing the charter of the First Bank of the United States in 1811 was that it threatened state banks and infringed upon states' rights. This notion of federal encroachment on an area generally considered to be the province of the states persisted throughout the 1820s and culminated in Andrew Jackson's veto of the recharter of the Second Bank of the United States in 1832.[20] President Jackson, noting that the states overwhelmingly opposed the recharter, called the Bank "subversive of the rights of the States"[21] and used his veto message as an occasion for recalling the importance of state sovereignty:

> Nor is our Government to be maintained or our Union preserved by invasions of the rights and powers of the several States. In thus attempting to make our General Government strong we make it weak. Its true strength consists in leaving individuals and States as much as possible to themselves—in making itself felt, not in its power, but in its beneficence; not in its control, but in its protection; not in binding the States more closely to the center, but leaving each to move unobstructed in its proper orbit.[22]

The perceived need to bear in mind states' rights when formulating bank policy, as articulated here by one of the foremost advocates of the state sovereignty principle, has endured as one of the central themes in financial institutions regulation. Moreover, as will be shown, the Jacksonian brand of federalism—and for that matter the Jacksonian fear of big banks—has profoundly influenced the scope of geographic restrictions on banking.[23]

The charter of the Second Bank of the United States expired in 1836, and for almost three decades thereafter, America had

---

[18] JONATHAN R. MACEY & GEOFFREY P. MILLER, BANKING LAW AND REGULATION 3, 5 (1992).

[19] The First Bank of the United States (1792–1811), headquartered in Philadelphia, had eight branches in other states. The Second Bank of the United States (1816–1836), also headquartered in Philadelphia, had 25 branches. *See A History of Interstate Banking in the U.S.*, BANKING POL'Y REP., Sept. 5, 1994, at 9.

[20] MACEY & MILLER, *supra* note 18, at 6, 8.

[21] Message by President Andrew Jackson Vetoing the Bank Recharter, July 10, 1832, *reprinted in* SYMONS & WHITE, *supra* note 15, at 14.

[22] *Id.* at 20.

[23] *See infra* part III.

neither a central bank nor any federally chartered banks. The states retained all chartering and supervisory responsibilities. State law governed the extent to which state banks could branch. Although state statutes flatly prohibiting branching were rare, in practice there was no branching in the New England states and limited branching in the South; extensive branching existed only in the western (currently midwestern) states.[24]

With the passage of the National Bank Act in 1864,[25] it became possible for the first time to open a bank under a national charter. The "dual banking system," in which state and national banks coexisted in the same markets, thus emerged. Competition developed between the two types of chartering authorities, both of which sought to attract banking organizations, along with the revenues and employment opportunities that flowed from their activities.[26] National banks were required to operate out of a single office; they were not allowed to branch.[27] After the Civil War, the branch systems of state-chartered banks also withered. Many state banks converted to national charters, and those that remained tended to conduct their affairs from a single office or, at most, a small number of offices located near one another. Thus, branching was relatively rare in nineteenth-century America.

This "unit banking system"—a system of widely dispersed banks, each doing business in a single location—resulted not from government-imposed restrictions but from market forces and community convictions. The highly localized economy created little need for long-distance banking services.[28] The unit banking system was appropriate in light of the economic conditions, and remained the dominant model in American banking for half a century.[29]

In addition, this system was based on an ideology that was fundamental to the American world view. Perhaps because the availability of credit was essential to the cultivation of the entrepreneurial spirit and the fulfillment of the American dream, community banks came to view themselves as performing an

---

[24] MACEY & MILLER, *supra* note 18, at 9, 13.

[25] Act of June 3, 1864, ch. 106, 13 Stat. 99 (codified in scattered sections of 12 U.S.C. (1994)).

[26] SYMONS & WHITE, *supra* note 15, at 24.

[27] MACEY & MILLER, *supra* note 18, at 13. *See also* GAO 1993, *supra* note 10, at 130.

[28] SYMONS & WHITE, *supra* note 15, at 97; Wilmarth, *supra* note 9, at 972.

[29] MACEY & MILLER, *supra* note 18, at 11.

essential function. They portrayed themselves as uniquely qualified to serve the credit needs of their communities, in a way that no larger, more far-reaching banking organization possibly could. One tenet of the small banks' "local" philosophy was that national banks should not be allowed to branch. At an American Bankers Association convention in 1902, for example, a group of community bankers argued that branching would

> create a brood of two hundred or three hundred great central banks, with 10,000 to 15,000 branches in large cities as well as small, and as such branches would have no capital and only figure-head management, individualism in management would cease, local taxes [would] be evaded, [there would be] no home distribution of profits, local progress [would be] retarded, in short, the great central banks would skim the cream from the whole country to enrich [their own] exchequers.[30]

Thus, at the turn of the century, community bankers were laying the foundations for the arguments that they would deploy in all of the great interstate banking debates of the twentieth century, as the threat posed by the big banks steadily intensified.[31]

B. *The Early to Middle Twentieth Century: Branching, Consolidation, and Resulting Limitations*

Eventually the unit banking system yielded to a new model. In the 1910s and 1920s, banks grew larger and expanded farther than they had previously. Independent banks came under common ownership and control, resulting in sprawling networks of interrelated, though legally separate, banking offices. Consolidation was accomplished in several different ways: chain banking (separately chartered banks owned by a single individual), group banking (separately chartered banks owned by a BHC), and merger (acquisition of one bank by another).[32]

Concurrent with this wave of consolidation and affiliation, states began to enact statutes permitting state-chartered banks to branch in-state.[33] Branching caught on quickly: the number of

---

[30] Community bankers' remarks at the 1902 American Bankers Association convention, *reprinted in* MACEY & MILLER, *supra* note 18, at 14.

[31] *See infra* part III.

[32] MACEY & MILLER, *supra* note 18, at 17, 19–20; Geoffrey P. Miller, *Legal Restrictions on Bank Consolidation: An Economic Analysis*, 77 IOWA L. REV. 1083, 1091–92 (1992).

[33] Among the first were California, Delaware, Florida, Georgia, Maine, Massachu-

branch banks in the United States nearly tripled from 1920 to 1930.[34] For a time it appeared that national banks would be excluded from the branching boom. The Supreme Court held in 1924 that national banks had no general power to branch, absent explicit authorization by Congress.[35] To remedy this imbalance and establish "competitive equality"[36] between state and national banks, Congress enacted the McFadden Act in 1927, pursuant to which national banks could branch to the same extent as state banks.[37] Thus, by the late 1920s, the American banking landscape had been radically altered by the dual phenomena of branching and consolidation.

The extent of bank expansion should not, however, be exaggerated. The McFadden Act is as noteworthy for what it failed to do as for what it actually did. It was the first of two major instances in which Congress declined to take the lead in authorizing bank expansion. Rather than directly empowering national banks to branch and deciding on a principled basis how far such branching should extend, Congress compromised by referring national banks to the relevant state statutes.[38] In short, Congress left state legislatures to be the pioneers. Few states, however, were enthusiastic about liberalizing their branching rules. In the 1920s and 1930s, only a minority of states permitted branching, and those that did often provided for branching only within the bank's home city or county, rather than statewide.[39] Branching across state lines was not allowed under any circumstances:[40] no

---

setts, and New York. MACEY & MILLER, *supra* note 18, at 18; Wilmarth, *supra* note 9, at 1092.

[34] In 1920 there were 1281; in 1930 there were 3522. MACEY & MILLER, *supra* note 18, at 18.

[35] First Nat'l Bank in St. Louis v. Missouri, 263 U.S. 640 (1924).

[36] This phrase, originally Representative McFadden's, was adopted and emphasized by the Supreme Court in First Nat'l Bank of Logan v. Walker Bank & Trust Co., 385 U.S. 252 (1966).

[37] Initially a national bank could only branch to the same extent as state banks within its *city of residence*. In 1933, the law was amended to allow for national bank branching to the same extent as state banks *throughout the state*. 12 U.S.C. § 36 (1994).

[38] *See* H. Rodgin Cohen, *Interstate Banking: Myth and Reality*, 18 LOY. L.A. L. REV. 965, 967 (1985) [hereinafter Cohen 1985] ("The compromise decision to confine branching expansion to state borders was unaccompanied by any detailed analysis of why the state, rather than a larger or smaller geographic area, was appropriate. There was no indication that Congress regarded the state as an economic unit. Indeed, an historical analysis would demonstrate that state boundary lines were not drawn for such purpose, much less to follow the demands of banking customers.").

[39] *See* CARTER H. GOLEMBE & DAVID S. HOLLAND, FEDERAL REGULATION OF BANKING 1983-84 118–21 (1983).

[40] The one exception was that foreign banks operating in the United States were allowed to branch interstate. MACEY & MILLER, *supra* note 18, at 26. Congress put an

state authorized it for its state-chartered banks, and at the federal level, the McFadden Act was construed as prohibiting (by negative implication) interstate branching by national banks.[41]

Thus, the law remained restrictive—so restrictive that it began to lag behind the economic reality. In an agrarian economy with limited means of transportation and communication, there had been little need or desire for permissive rules concerning bank expansion.[42] But as a result of technological advances and the increasing mobility of the population, the market was gradually starting to call for more branch banking than the regulatory framework allowed. A prominent big-city banker warned Congress in 1930:

> There is only one way to do banking, and that is on a branch basis, with one capitalization, one charter, and one responsibility . . . . It is coming, gentlemen, and you cannot stop it, and you are bucking up against a stone wall if you try. You cannot buck economic forces.[43]

Two years later, a bill to allow national banks to branch statewide, regardless of state law, was defeated at the hands of the powerful community bankers' lobby.[44] Thus, although market principles had begun to militate in favor of loosening geographic constraints, Congress chose not to facilitate branching any further.

The second instance in which Congress effectively impeded the progress of bank expansion by deferring to the states was when it enacted the Douglas Amendment in 1956. The Douglas Amendment was Congress's response to an increasingly popular industry technique for circumventing geographic restrictions—the BHC. To achieve the effect of interstate branching without violating the prohibition against it, a banking organization could simply use the BHC format to cross state lines. It had long been the rule that a BHC could acquire banks in more than one state

---

end to this practice by enacting the International Banking Act of 1978, which prohibited foreign banks from branching outside their home state. 12 U.S.C. § 3103(a) (1994).

[41] *See* SYMONS & WHITE, *supra* note 15, at 90.

[42] *See* Joann Senzel Nestor, *Interstate Branch Banking Reform: Preserving the Policies Underlying the McFadden Act*, 72 B.U. L. REV. 607, 614 (1992); Carl Felsenfeld, *Electronic Banking and its Effects on Interstate Branching Restrictions—An Analytic Approach*, 54 FORDHAM L. REV. 1019, 1059 (1986); SYMONS & WHITE, *supra* note 15, at 97.

[43] *Branch, Chain, and Group Banking: Hearings on H.R. 141 Before the House Comm. on Banking and Currency*, 71st Cong., 2d Sess. 1537 (1930) (testimony of A.P. Giannini, founder of Transamerica Corp).

[44] Wilmarth, *supra* note 9, at 973–74.

and, so long as it held them as subsidiaries, legally independent of one another, operate them under a common management strategy. The BHC format first became popular in the 1920s,[45] and by the 1950s, use of the BHC as a device for interstate expansion had become so widespread that community bankers began aggressively to seek protection against the competitive threat that the multistate BHCs posed. Congress responded to their pleas by enacting the BHCA, whose legislative intent the Supreme Court described as follows:

> One of the major purposes of the BHCA was to eliminate this loophole. As enacted by the House in 1955, the BHCA contained a flat ban on interstate bank acquisitions. The legislative history from the House makes it clear that the policies of community control and local responsiveness of banks inspired this flat ban.[46]

The community bankers' victory was tempered somewhat by the addition, on the Senate side, of the Douglas Amendment, which allowed individual states to lift the ban at their discretion.[47] Although the legislation as enacted was more permissive than the original House version, it was hardly a generous grant of authority to bank interstate. As the Supreme Court noted, even with the addition of the Douglas Amendment, the BHCA was still essentially restrictive: "[T]he broader purposes underlying the BHCA as a whole and the Douglas Amendment in particular [were] to retain local, community-based control over banking."[48] The Douglas Amendment was the McFadden Act of interstate banking: Congress had once again ceded to the state legislatures.

C. *The 1980s and 1990s: Increasing Tension Surrounding Interstate Banking*

It was a long time before any state exercised its option under the Douglas Amendment to lift the ban on interstate bank acquisition. Until Maine adopted legislation permitting entry by out-of-state BHCs in 1975,[49] followed by additional states in the 1980s, the Douglas Amendment effectively functioned as a bar

---

[45] MACEY & MILLER, *supra* note 18, at 20.

[46] Northeast Bancorp v. Board of Governors of the Fed. Reserve Sys., 472 U.S. 159, 169 (1985).

[47] *See supra* noted 7–8 and accompanying text.

[48] *Northeast Bancorp*, 472 U.S. at 172.

[49] *See* Wilmarth, *supra* note 9, at 977.

to interstate banking. Even when the states elected to loosen the constraints on interstate expansion, they often did so in a limited way: they authorized entry by BHCs from certain specified states rather than from all states. For example, in 1982 Massachusetts enacted a statute permitting acquisition of Massachusetts-based banks by BHCs from New England states only; Connecticut did the same the following year.[50] This technique of partially lifting the Douglas Amendment ban, so as to authorize only regional interstate banking,[51] perpetuated protectionism and geographic restrictions. After the Supreme Court approved this practice in 1985,[52] the number of states permitting at least limited entry by out-of-state BHCs increased dramatically. As of 1993, every state except Hawaii had an interstate banking statute. Although about a third of the states retained some version of the regional formula, fully two-thirds (34 states) permitted entry by BHCs from any state in the nation.[53] At the same time, the substantial liberalization of intrastate branching rules has enabled banks to broaden their fields of operation. As of 1994, full statewide branching was the rule in thirty-seven states, compared to sixteen in 1960.[54] Thus, on the eve of the Riegle-Neal Act's passage, the rules circumscribing bank operations were at their most lenient—on account of actions taken at the state rather than the federal level.[55]

The barriers to bank expansion had not, however, been completely dismantled. On the eve of passage of the Riegle-Neal Act, restrictions on intrastate branching and interstate acquisitions by BHCs remained in effect in a significant number of jurisdictions (thirteen and sixteen states, respectively).[56] As for

---

[50] *Northeast Bancorp*, 472 U.S. at 164.

[51] These region-specific statutes gave rise to the term "interstate compact" to describe the practice of admitting BHCs from some states but not others. There were no actual "compacts," however, because the states did not actually meet and agree on reciprocal arrangements. Rather, each state acted independently, listing the states whose BHCs it would grant preferential treatment. Donald T. Savage, *Interstate Banking: A Status Report*, 79 Fed. Res. Bull. 12,1075 at 31 (Dec. 1993). In New England and the Southeast, the states' lists more or less matched; hence it was natural that major BHCs like Fleet and NationsBank expanded throughout these regions. By contrast, the midwestern states had different conceptions of which states constituted the midwest, with the result that there were no distinctly midwestern BHCs along the lines of Fleet in the northeast or NationsBank in the south. JOHN B. MCCOY ET AL., BOTTOMLINE BANKING: MEETING THE CHALLENGES FOR SURVIVAL & SUCCESS 143 (1994).

[52] *Northeast Bancorp*, 472 U.S. 159.

[53] GAO 1993, *supra* note 10, at 21.

[54] *Id.* at 15; MCCOY ET AL., *supra* note 51, at 63.

[55] *See* Robert Litan, *Interstate Banking and Product-Line Freedom*, 9 YALE J. ON REG. 521, 529 (1992).

[56] *See supra* notes 53–54 and accompanying text.

interstate branching, it was and still is virtually nonexistent. Theoretically it is possible for state-chartered Federal Reserve nonmember banks[57] to branch across state lines, and a few states allow such branching. They are a small minority, however, and as a practical matter banks have not yet begun to open branches out of state.[58]

Congress's repeated rejection of proposals to eliminate geographic constraints indicates its determination to leave at least some of these constraints in place, and shows the persistence of tension between community desires for local control and market pressures for more expansion. The first of several failed attempts at reform occurred in 1981, when Congress took no action in response to a Treasury Department report calling for more interstate banking.[59] A House Committee bill in 1985 was also unsuccessful.[60] In 1987, the ongoing struggle was highlighted by the controversy surrounding a proposed exception to the Douglas Amendment. That year, Congress passed a measure that trumped the Douglas Amendment by allowing interstate acquisition of failed or failing banks irrespective of state law.[61] Even this slight relaxation of the Douglas Amendment, however, undertaken in the face of a serious emergency in the banking industry, met with strong opposition from advocates of states' rights and community bankers. A contemporary news account noted:

> [The] bill is under fire from consumer groups, who question its need and decry its lack of consumer and anticoncentration protections . . . , and by state bank supervisors and others who believe its effect would inhibit in-state acquisitions of failing banks and bank holding companies . . . in favor of large out-of-state, expansion-minded bank holding companies eager to enter new markets.
>
> Many of the bill's critics have said that they are not convinced that the troubled bank situation is severe enough to warrant changes to the Douglas Amendment . . . .
>
> Independent bankers . . . are concerned that [certain pro-

---

[57] State-chartered banks that belong to the Federal Reserve System are called "member banks"; those that do not are called "nonmember banks."

[58] *See* GENERAL ACCOUNTING OFFICE, INTERSTATE BANKING: EXPERIENCES IN THREE WESTERN STATES, GAO/GGD-95-35, Dec. 1994, at 3 [hereinafter GAO 1994].

[59] U.S. DEP'T OF THE TREASURY, GEOGRAPHIC RESTRICTIONS ON COMMERCIAL BANKING IN THE UNITED STATES (1981).

[60] Depository Institutions Acquisitions Act of 1985, H.R. 2707, 99th Cong., 1st Sess. (1985). *See* Wilmarth, *supra* note 9, at 976–77.

[61] Competitive Equality Banking Act of 1987 (CEBA), Pub. L. No. 100-86, 101 Stat. 552, codified at 12 U.S.C. § 1823(f)(4) (1994).

> visions of the bill] would open the door wide to interstate banking by large holding companies.[62]

As this account demonstrates, the Jeffersonian and Jacksonian fear of concentrated banking power and distrust of banks touched with what Louis Brandeis called "the curse of bigness"[63] were still very much alive in the 1980s. The independent banks had come to view themselves as having a moral mission: to support their local communities and, in a truly populist sense, stand by the common man. Proponents of the "local" banking model asserted that the "money-center banks"[64] were seeking to obtain a stranglehold over banking markets everywhere and to siphon deposits out of them. Seizing upon the stereotype of the avaricious out-of-town banker, the community banks championed the cause of local control.

Indeed, the Supreme Court has endorsed the belief of policymakers and the public that banks are closely tied to their communities, and that banks should be regulated and supervised on a local level. Two landmark opinions reflect the belief that banking is fundamentally a local matter. The first concerns the application of antitrust laws to banking. The Clayton Act, in prohibiting mergers "where in any line of commerce *in any section of the country*, the effect of such acquisition may be substantially to lessen competition,"[65] calls for an inquiry into the bounds of a bank's geographic market. In *United States v. Philadelphia National Bank*, the Court held that a bank's geographic market for antitrust purposes is the local area only. In determining this market, the Court reflected on the dimensions of the banking area:

> Individuals and corporations typically confer the bulk of their patronage on banks in their local community; they find it impractical to conduct their banking business at a distance

---

[62] *Regulators Defend Failing Bank Bill, Texas Bank Commissioner Questions Scope*, WASH. FIN. REP., May 12, 1986, at 784.

[63] LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 162 (1914).

[64] The money-center banks are large banks headquartered in the nation's top financial centers, such as Citibank, First Chicago, and BankAmerica. The term "money-center bank" is often used pejoratively. The mention of Citibank, for example, may be intended to alarm or inflame. *See, e.g., Costs and Benefits of Interstate Banking and Branching: Hearings Before the Subcomm. on Financial Institutions Supervision, Regulation and Deposit Insurance of the House Comm. on Banking, Finance and Urban Affairs*, 103d Cong., 1st Sess. 31 (1993) [hereinafter *June 1993 Hearing*] (testimony of Herman D. Farrell, chairman of the Bank Committee of the New York State Assembly) ("Whenever you are frightened of anything, you mention Citibank.").

[65] 15 U.S.C. § 7 (1994) (emphasis added).

> . . . .
>
> Large borrowers and large depositors, the record shows, may find it practical to do a large part of their banking business outside their home community; very small borrowers and depositors may, as a practical matter, be confined to bank offices in their immediate neighborhood; and customers of intermediate size, it would appear, deal with banks within an area intermediate between these extremes . . . . So also, some banking services are evidently more local in nature than others. But that in banking the relevant geographic market is a function of each separate customer's economic scale means simply that a workable compromise must be found . . . . We think that the four-county Philadelphia metropolitan area, which state law apparently recognizes as a meaningful banking community in allowing Philadelphia banks to branch within it, and which would seem roughly to delineate the area in which bank customers that are neither very large nor very small find it practical to do their banking business, is a more appropriate "section of the country" in which to appraise the instant merger than any larger or smaller or different area.[66]

Thus, the Court acknowledged the difficulty of the question but nevertheless adopted the traditional view that banking should be seen as essentially a local matter. Later, in a famous phrase in *Lewis v. BT Investment Managers, Inc.*,[67] the Court made the point more directly: "We readily accept the submission that, both as a matter of history and as a matter of present commercial reality, banking and related financial activities are of profound local concern."[68] The question, however, is whether by the time the Court articulated this enduring conception of banking—not in 1880 but in 1980—it had outlived its usefulness.

In 1991, Congress came close to recognizing the obsolescence of geographic restrictions when it considered the proposed Financial Institutions Safety and Consumer Choice Act (FISCCA),[69] which would have repealed the McFadden Act and the Douglas Amendment.[70] FISCCA had the support of the Bush administration and the Treasury Department, which the same year issued an influential report advocating nationwide banking and branch-

---

[66] United States v. Philadelphia Nat'l Bank, 374 U.S. 321, 358–61 (1963).

[67] 447 U.S. 27 (1980).

[68] *Id.* at 30.

[69] Financial Institutions Safety and Consumer Choice Act of 1991 (FISCCA), S. 713, 102d Cong., 1st Sess. (1991), H.R. 1505, 102d Cong., 1st Sess. (1991).

[70] *See* Jonathan R. Macey & Geoffrey P. Miller, *America's Banking System*, 69 WASH. U. L.Q. 769, 793 (1991).

ing.[71] But instead of making major modifications to the regulatory structure, Congress confined its attention to enforcement issues and the recapitalization of the Bank Insurance Fund (BIF). To the disappointment of those who had advocated transition to the "global" banking model,[72] the Federal Deposit Insurance Corporation Improvement Act of 1991 (FDICIA) did not address the question of geography.[73] The redrawing of boundaries for the banking industry had once again been postponed. Indeed, when Congress began to study the issue in earnest in 1993, the arguments on both sides were already familiar, and there was a sense that it was time to take action at last.

II. CIRCUMVENTION OF GEOGRAPHIC RESTRICTIONS

A. *The Reality of Nationwide Banking*

The state and federal legislation explored in Part I of this Article only matters up to a point. Banking is a business, and shrewd businesspeople, together with their lawyers, inevitably find ways to circumvent the rules and conduct their affairs as they see fit.[74] No study of the geographic reach of banks would be complete without an analysis of the many devices that banks have employed to carry out operations in multiple regions without violating the law. As a result of these innovative techniques, much of the inconvenience to consumers that might have resulted from strict adherence to both the letter and the spirit of the geographic restrictions has been avoided. Not all of the

---

[71] U.S. DEP'T OF THE TREASURY, MODERNIZING THE FINANCIAL SYSTEM: RECOMMENDATIONS FOR SAFER, MORE COMPETITIVE BANKS (1991) [hereinafter MODERNIZING THE FINANCIAL SYSTEM].

[72] *See, e.g., Impact of Bank Reform Proposals on Consumers: Hearing Before the Subcomm. on Consumer Affairs and Coinage of the Comm. on Banking, Finance and Urban Affairs, House of Representatives*, 102d Cong., 1st Sess. 8 (1991) [hereinafter *Apr. 1991 Hearing*] (testimony of Jay Powell, Assistant Secretary for Domestic Finance, U.S. Dep't of the Treasury) ("[C]omprehensive reform, and not piecemeal reform, is what is needed. If we only tinker with the problem—for example, by simply recapitalizing the Bank Insurance Fund—then we will not have addressed the underlying causes that have brought the Fund to its present state."); Alfred J.T. Byrne & Martha L. Coulter, *Safety and Soundness in Banking Reform*, 69 WASH. U. L.Q. 679, 685 (1991).

[73] Federal Deposit Insurance Corporation Act of 1991, Pub. L. No. 102-242, 105 Stat. 2236 (1991) (codified as amended in scattered sections of 12 U.S.C.); *see* Litan, *supra* note 55, at 522.

[74] *See* Geoffrey P. Miller, *Interstate Banking in the Court*, 1985 SUP. CT. REV. 179, 183 (1985) ("In banking law, . . . the market often mocks the laws . . . . Banks found any number of ways to do business interstate.").

banks' stratagems have been extra-legal; some have been explicitly sanctioned by the judiciary, or depend upon a certain operation's falling just outside the applicable statutory definition. Others are simply part of commonly accepted practice.

Given these loopholes, it can be said without exaggeration that nationwide banking was already a reality when Congress passed the Riegle-Neal Act.[75] Most Americans are probably largely unaware of the limitations that have traditionally been imposed upon banks. One congressman noted during floor debates on the Riegle-Neal Act: "I am sure that the vast majority of our constituents have no idea that there is a Federal prohibition against interstate banking."[76] If this is the public perception, then the story told in Part I, which emphasized the barriers to nationwide banking, needs to be supplemented with an examination of those aspects of the banking business that have not been confined to narrow zones. It should be borne in mind throughout this discussion that the relative ease with which Americans can obtain various forms of long-distance banking services is an achievement not of government, but of the marketplace.[77]

B. *Interstate Expansion via the BHC*

Interstate expansion using the BHC, as described above,[78] has proven a useful tool for fashioning large, multistate banking organizations that are the functional equivalent of branching networks. BHCs have become the dominant banking structure in America: as of 1994 they held about ninety-four percent of all U.S. banking assets, and all but one of the fifty largest banks in the country were part of a BHC system.[79] Using the BHC device, a single banking enterprise—with one management, one philosophy, and one business strategy—can enter markets in multiple

---

[75] *See, e.g.*, Miller, *supra* note 32, at 1084.

[76] 140 CONG. REC. H6774, 6779 (daily ed. Aug. 4, 1994) (statement of Representative Hoyer (D-Md.)).

[77] Many commentators advocate letting the market, rather than regulation, determine the size and shape of the banking area. *See, e.g.*, Miller, *supra* note 32, at 1131 ("It is market forces, together with technological innovation, that are driving the U.S. banking industry in the direction of ever-increasing geographic expansion . . . . Markets, not politicians or bureaucrats, should decide the future structure of the American banking industry.").

[78] *See supra* notes 45–48 and accompanying text.

[79] GAO 1993, *supra* note 10, at 16; H. Rodgin Cohen, *Will Interstate Branching Overwhelm Old Advantages of Bank Holding Companies?*, BANKING POL'Y REP., Aug. 15, 1994, at 9.

states and behave almost as if the walls dividing its separately incorporated bank subsidiaries did not exist. The subsidiaries can bear substantially the same name: for example, BankAmerica Corporation (a BHC headquartered in California) operates subsidiaries called Bank of America Arizona (a national bank with branches in Arizona) and Bank of America Texas (a national bank with branches in Texas).[80] The same "BankAmerica" logo is displayed on bank buildings in both states. As a result, a handful of BHC names are household words not just in one or two cities, but throughout entire regions of the country: for example, as of 1991, BancOne operated in fourteen states, First Interstate in thirteen, NationsBank in eleven, BankAmerica in ten, and Citicorp in seven.[81] Indeed, most large BHCs own banks in more than one state.[82]

From the consumer's perspective, the BHC arrangement is scarcely distinguishable from a single legal entity having branches in multiple states. The only limitation on a BHC's power to create the illusion of interstate branching is that the resemblance must not be too close. If the nominally independent banks act in substance like branches, particularly in the accommodation services they offer one another, the structure risks being deemed a branching network and thus voided on the authority of the McFadden Act.[83] Funds can, however, move relatively freely between sibling banks within a BHC family. Deposits collected by Bank *A* may be transferred or sold to affiliated Bank *B* using such devices as interbank deposits or the federal funds market.[84] In addition, customers of Bank *A* may rely on Bank *B* (located in another state) for certain types of banking services, using wire transfers or other forms of electronic fund transfers.[85] Many trans-

---

[80] Telephone Interview with Peter Magnani, customer service representative, BankAmerica Corporation, San Francisco, Cal. (Nov. 15, 1995).

[81] Janet L. Fix, *Banks Extend Their Reach: Deregulation Lets Banks Branch Out*, USA TODAY, Aug. 8, 1994, at B1.

[82] MACEY & MILLER, *supra* note 18, at 430.

[83] Accommodation services may not be substantially more extensive than traditional correspondent banking services. *See, e.g.*, Michigan Nat'l Corp., 64 Fed. Res. Bull. 127 (1978) (holding that "accommodation transaction services" program, whereby customers of one bank subsidiary could make deposits and withdrawals at another, constituted branching in substance and competitive impact); United States v. Citizens & S. Nat'l Bank, 422 U.S. 86 (1975) (holding that commonly owned banks were *de facto* branches).

[84] *See* GAO 1993, *supra* note 10, at 17 n.2.

[85] EFTs are computerized transfers among banks, involving no paper or intermediate steps such as the exchange of checks or the use of credit cards. *See* SYMONS & WHITE,

actions between affiliates that would otherwise be prohibited under section 23A of the Federal Reserve Act are permitted where both affiliates are banks; in other words, the restrictions are less stringent than in the case of dealings between a bank and an affiliated nonbank.[86] To be sure, bank affiliates cannot do for one another (or for one another's customers) what branches can do. There are no restrictions on the transfer of assets and liabilities among branches, which are all part of a single entity for legal and accounting purposes. Furthermore, bank subsidiaries, unlike branches, may not accept deposits from customers of affiliated banks.[87] Nonetheless, the distinction is sufficiently subtle that, practically speaking, widespread use of the BHC format has already created a nationwide banking system.

C. *Cooperation Between Banks*

Even banks that do not belong to a common BHC family may perform certain functions for one another, thus enabling customers to obtain services from banks not their own, often in other parts of the country. Correspondent banking, whereby one bank provides services in another's behalf in exchange for a fee, has long been an important feature of the American banking system.[88] Within the limits of the *Michigan National* doctrine,[89] a bank whose customers require out-of-state services can usually arrange for a correspondent bank to provide them. More generally, banks have traditionally cooperated—using transaction-specific contracts, joint ventures, and other forms of informal affiliation—in order to meet their customers' interstate banking needs.[90] These techniques allow coordination of efforts without necessitating common ownership, which would trigger the BHC rules. A loosely linked banking chain (where the common owner holds less than twenty-five percent of at least one of them, thereby

---

supra note 15, at 129. For a discussion of EFT services among affiliated banks, see Wilmarth, *supra* note 9, at 1016 n.272.

[86] 12 U.S.C. § 371c(d)(1) (1994).

[87] *See* GAO 1993, *supra* note 10, at 101–02. *See infra* notes 117–126 and accompanying text.

[88] Felsenfeld, *supra* note 42, at 1037.

[89] *See supra* note 83.

[90] *See* Peter C. Carstensen, *Public Policy Toward Interstate Bank Mergers*, 49 OHIO ST. L.J. 1397, 1408, 1413–14 (1989).

failing the BHC test)[91] is free to disregard the Douglas Amendment and engage in ad hoc interstate banking.[92]

D. *Activities Not Constituting "Branching"*

In addition, a variety of banking services are conducted across state lines by the simple device of falling outside the definition of "branch." The McFadden Act only prohibits interstate branching; activities that do not constitute branching therefore escape its operation. Pursuant to 12 U.S.C. § 36(f), a banking office rises to the level of a branch if "deposits are received, or checks paid, or money lent" on its premises. This definition has been the subject of much litigation, in which courts have generally construed the term "branch" broadly. The "deposits received" prong of the definition, for example, has a hair-trigger mechanism: even where the remote facility is an armored car that acts strictly as the customer's agent, so that funds are not formally considered "received" until deposited at the actual bank, the facility still constitutes a branch under 12 U.S.C. § 36(f).[93] Similarly, "customer-bank communication terminals" (CBCTs), including automatic teller machines (ATMs), have been held to constitute branches.[94] Even a trust office, which appears not to perform any of the three functions listed in 12 U.S.C. § 36(f), qualifies as a branch.[95] Many commentators have expressed concern that this definition of the term "branch" may be overly expansive.[96] Clearly, the propensity of courts to see a branch wherever there is a banking operation of any kind—even a machine or a truck—complicates the task of banks that, for the purpose of avoiding the McFadden Act, seek to fall outside of the definition.

Clever exploitation of the technicalities, however, has allowed banks to avoid the branch designation. A regulation promulgated by the Office of the Comptroller of the Currency (OCC) provides

---

[91] 12 U.S.C. § 1841(a)(2)(A) (1994). *See supra* note 6.

[92] Lawrence G. Goldberg, *What We Can Expect from Interstate Banking*, 12 J. BANKING & FIN. 51, 53 (1988).

[93] First Nat'l Bank in Plant City, Florida v. Dickinson, 396 U.S. 122 (1969).

[94] Independent Bankers Ass'n of America v. Smith, 534 F.2d 921 (D.C. Cir. 1976), *cert. denied*, 429 U.S. 862 (1976).

[95] St. Louis County Nat'l Bank v. Mercantile Trust Co. Nat'l Ass'n, 548 F.2d 716 (8th Cir. 1976), *cert. denied*, 433 U.S. 909 (1977).

[96] *See, e.g.*, David F. Freeman, *Interstate Banking Restrictions under the McFadden Act*, 72 VA. L. REV. 1119, 1122–23 (1986).

that CBCTs are machines "established (i.e., owned or rented)" by a national bank.[97] Hence the solution is for national banks to use ATMs without owning or renting them. The Second Circuit has held that an ATM owned, operated, and serviced by a grocery chain, and bearing only the grocer's logo, may be used by a national bank's customers to make deposits, withdrawals, transfers, and balance inquiries, without the ATM's being deemed a branch of the bank.[98] Similarly, rather than involving a merchant, Bank *A* may arrange to use Bank *B*'s ATM, which *B* owns or rents in a place where it can lawfully maintain a branch. Since the ATM is a branch to *B* but not to *A*, *A* may use it irrespective of geographic restrictions.[99]

The "own or rent" loophole has revolutionized American banking. At least with respect to basic deposit and withdrawal functions, ATMs make geographic boundaries irrelevant. Their popularity has increased rapidly: essentially absent from the landscape twenty years ago, they are now the preferred banking method of younger consumers.[100] To attract and retain increasingly mobile customers, banks now feel competitive pressure to associate themselves with the largest ATM networks, such as Cirrus, Plus, Yankee, and the New York Cash Exchange (NYCE).[101] These sprawling ATM networks are probably the principal reason that so many consumers are unaware of, and largely unaffected by, the geographic restrictions imposed on banks. Bank customers who need cash while traveling can easily access their accounts from machines located in airports, train stations, supermarkets, shopping centers, and banks all over the country and increasingly around the world.[102]

---

[97] "A 'CBCT branch' is an automated device, established (i.e., owned or rented) by a national bank at a location separate from the main office or a domestic branch, that: (1) Takes deposits, or (2) Disburses cash drawn against: (i) A customer's deposit account, or (ii) A customer's pre-approved loan account." 12 C.F.R. § 5.31(b) (1995).

[98] Independent Bankers Ass'n of New York State, Inc. v. Marine Midland Bank, N.A., 757 F.2d 453 (2d Cir. 1985), *cert. denied*, 476 U.S. 1186 (1986). This reasoning has been extended to point of sale (POS) terminals as well. A POS terminal is a machine allowing shoppers to use a debit card to pay for purchases at the register of a supermarket or other retail store. The terminal is directly linked to a bank, so that the amount of the sale can be immediately transferred from the customer's account to the merchant's account. *See* Nestor, *supra* note 42, at 617 n.82; James C. Sivon, *New Law May Warrant Search for Alternatives to Interstate Branching*, BANKING POL'Y REP., Oct. 17, 1994, at 5.

[99] *See* Felsenfeld, *supra* note 42, at 1036–37.

[100] McCOY ET AL., *supra* note 51, 231–32.

[101] Paul Gibson, *Betting the Branch*, INSTITUTIONAL INVESTOR, Sept. 1994, at 155.

[102] *See* Savage, *supra* note 51, at 1081.

To be sure, ATMs do not represent a comprehensive solution to the problem of interstate banking. They can only accommodate certain limited types of transactions. To the extent that face-to-face consultations and counseling are necessary—or simply desirable, as studies show they are to many consumers, particularly those middle-aged and older[103]—ATMs are inadequate.[104] During the drafting of the 1994 legislation, Congressmen Neal and McCollum argued that the shortcomings of ATMs were a compelling reason to permit interstate branching:

> ATMs are not substitutes for full-service brick-and-mortar branches. Customers at ATMs face per day withdrawal limits, and ATMs are generally unable to cash a third-party check, or dispense travelers, cashiers', or personal checks. They cannot assist customers in opening new accounts, purchasing certificates of deposit or mutual funds, or providing advice about other products and services.[105]

In addition, using an ATM not immediately affiliated with the account bank may entail a surcharge, which lessens the ATM's usefulness as a means of conducting banking business at a distance. In light of the convenience that ATMs afford, however, particularly in the area of ready access to cash, they represent a major step toward instituting nationwide banking within the framework of the otherwise balkanized American banking system.

Other types of facilities, in addition to self-serve banking machines, may also fall short of the definition of "branch" and thus be permitted to operate outside the usual geographic bounds. Offices devoted exclusively to marketing and promotional activities, for example, are exempt from geographic restrictions so long as they are not used to conduct general banking business. A bank may maintain a so-called "representative office" out of state for the purpose of distributing information, developing contacts, and promoting business. "Calling officers" may travel out of state to market banking services, usually to corporate and institutional clients, without risk that their activities will be deemed branching.[106]

---

[103] McCoy ET AL., *supra* note 51, at 224, 231–32.

[104] Travelers, however, probably do not require sophisticated personal service. *See* Wilmarth, *supra* note 9, at 1017 (explaining that travelers can "charge purchases or obtain cash advances by using nationally-recognized credit cards . . . [and] do not usually need the kind of full banking services provided by brick-and-mortar branches.").

[105] H.R. REP. No. 448, 103d Cong., 2d Sess. 59 (1994) [hereinafter Neal & McCollum] (additional views of Mr. Neal and Mr. McCollum).

[106] Freeman, *supra* note 96, at 1121 n.9; Cohen 1985, *supra* note 38, at 973.

Similarly, national banks are permitted to establish loan production offices (LPOs) throughout the country, for the purpose of soliciting borrowers and carrying out the preliminary steps associated with originating a loan (i.e., assembling credit information, preparing the application, etc.). The LPO must not, however, perform more than a ministerial function. All credit decisions and disbursements of funds must be made by personnel at the bank itself.[107] It should be emphasized that the LPO exception is based upon the LPO's confining itself to a narrow sector of the banking business. LPOs are for lending only; they may not attend to a customer's other banking needs. In reviewing OCC letters interpreting the LPO exception, the D.C. Circuit noted:

> These letters have permitted only interviewing, counseling, and assisting applicants on loan rates and terms; they expressly have forbidden executing notes and security agreements, making forms available for opening checking or savings accounts, counseling customers on other banking services, supplying existing customers with information on their deposit accounts, disbursing loan funds, and accepting loan payments.[108]

Despite these limitations, it is significant that LPOs are brick-and-mortar facilities, staffed by human beings who have face-to-face contact with customers. Consequently, they represent a form of interstate banking one stage more advanced than ATMs.

E. *Out-of-State Lending Activities*

As the LPO exception implies, banks are free to lend into any market they wish, irrespective of geographic considerations. A bank in one part of the country may hold loans in another (or, more commonly, loan participations)[109] without running afoul of the interstate rules. In practice, however, banks often hesitate to lend into markets where they do not have a physical presence for two principal reasons: it is harder to obtain adequate information on distant borrowers and economic climates for the pur-

---

[107] 12 C.F.R. § 7.7380(b) (1995); 12 C.F.R. § 250.141 (1995); *see also* Independent Bankers Ass'n of America v. Heimann, 627 F.2d 486 (D.C. Cir. 1980).

[108] *Heimann*, 627 F.2d at 488 n.**.

[109] Rather than holding an entire loan, a bank will often subscribe to a piece of the loan called a "participation." A bank that sells participations to other banks is said to have "participated out" its loans.

pose of evaluating creditworthiness, and the monitoring costs once credit is extended are higher for faraway borrowers.[110]

Nevertheless, long-distance lending does occur. For example, one form of lending that is conducted on a particularly wide interstate basis is consumer credit card lending. As a member of an interbank settlement organization such as Visa or MasterCard, a bank may issue credit cards to both depositors and nondepositors, irrespective of their state of residence.[111] The Supreme Court has held that national banks are free to extend credit to out-of-state customers and to issue credit cards to them for use wherever they wish: "Minnesota residents can . . . use their Omaha Bank [Nebraska] BankAmericards to purchase services in the State of New York or mail-order goods from the State of Michigan."[112] Many large banks either purchase credit card portfolios from banks located in other parts of the country, or solicit their own credit card customers nationwide using direct mailings and advertising.[113] Moreover, a bank may consider relocating its entire credit card operation to another state in order to benefit from more lenient regulations. For example, since South Dakota has no usury ceiling,[114] Citibank transferred its credit card business to its affiliate in South Dakota, Citibank (South Dakota), N.A. The result is that there is no limit on the interest rate Citibank charges its cardholders.[115] From its base in South Dakota, Citibank has pursued an ambitious marketing strategy from coast to coast and has become the nation's largest issuer of credit cards.[116] Thus, despite the geographic restrictions imposed upon banks, in practice the lending component of the banking business is far from parochial.

---

[110] GAO 1993, *supra* note 10, at 111.

[111] *See* Christopher C. DeMuth, *The Case Against Credit Card Interest Rate Regulation*, 3 YALE J. ON REG. 201, 205–07 (1986).

[112] Marquette Nat'l Bank of Minneapolis v. First of Omaha Service Corp., 439 U.S. 299, 312 (1978).

[113] Cohen 1985, *Interstate Banking*, *supra* note 38, at 973.

[114] Usury is the practice of charging an exorbitant rate of interest on a loan. Usury law is generally state law; different states have different interest rate caps. MACEY & MILLER, *supra* note 18, at 188.

[115] Citicorp, 67 Fed. Res. Bull. 181 (1981).

[116] *See* John W. Milligan, *Citicorp Changes Its Retail Prescription*, U.S. BANKER, July 1994, at 33; *Citicorp's Bag of Troubles*, ECONOMIST, Apr. 14, 1990, at 85.

F. *Out-of-State Deposit Collection*

Similarly, on the liabilities side of the balance sheet, banks may collect deposits from anywhere in the nation or abroad.[117] Their ability to attract deposits from a distance, however, is hampered by the restrictions on branching. One of the fundamental tenets of American banking regulation is that the deposit-taking function is sacred; it is what sets banks apart from other types of financial institutions.[118] Indeed, concerns about deposit-taking are at the heart of the McFadden Act's prohibition on interstate branching.[119] What the community banks are really saying when they defend the anti-branching rules is that they want to be insulated from competition in the deposit business. During drafting of the Riegle-Neal Act, the chairman of a small bank testified before a House committee:

> Why do the large financial institutions want this bill? If they can now make loans all over the country, which they can, if they can transfer money by wire all over the country, which they can, what is it they are after?
>
> And I think we have to look directly at what they are after; the core deposits in the small communities across the country. They are after a bigger share of a shrinking pie.[120]

Rather than face the threat that out-of-state banks will enter their markets and soak up deposits, the independent banks argue that they are entitled to a first claim on any funds that local residents wish to invest in savings accounts and certificates of deposit (CDs).[121] The McFadden Act partially protects this claim by narrowly circumscribing branch networks. An out-of-state bank will naturally have difficulty attracting local deposits if it cannot maintain a local branch. Nor can it employ the usual stratagems for establishing a local presence without branching: there is no depository analog to the LPO,[122] and customers tend to use off-site

---

[117] MACEY & MILLER, *supra* note 18, at 262.

[118] Other types of financial institutions include insurance companies, broker-dealers, and investment companies. *See infra* notes 143–145 and accompanying text.

[119] Felsenfeld, *supra* note 42, at 1047–49.

[120] *Interstate Banking and Branching: Hearing Before the Subcomm. on Financial Institutions Supervision, Regulation and Deposit Insurance of the House Comm. on Banking, Finance and Urban Affairs*, 103d Cong., 1st Sess. 15 (1993) [hereinafter *July 1993 Hearing*] (testimony of James R. Lauffer, chairman, president, and CEO, First National Bank of Herminie).

[121] A related concern is that these funds will then be funnelled out to borrowers in distant parts of the country. *See infra* notes 248–251 and accompanying text.

[122] Any facility purporting to be a mere "deposit production office" would likely be

ATMs far more for withdrawing cash than for making deposits.[123]

None of this is to suggest, however, that banks have not succeeded in extending their deposit operations across state lines. On the contrary, they are able to conduct a nationwide deposit business using two devices. The first is banking by mail and telephone: rather than stopping by the neighborhood branch, the customer transacts business from her home.[124] The second strategy—one that has had a profound impact on the industry—is to rely upon the services of professional deposit brokers, who act as matchmakers between depositors seeking high interest rates and banks seeking to raise funds.[125] Particularly where large institutional deposits are concerned, deposit brokering has given rise to a truly nationwide deposit market.[126]

G. *The Main Office Loophole*

Finally, a century-old loophole that had long gone unnoticed has been the focus of considerable attention recently, because it permits actual interstate branching in certain circumstances. Pursuant to 12 U.S.C. § 30(b), enacted in 1886, a national bank may relocate its main office anywhere within a thirty-mile radius.[127] This provision has been interpreted to include moves across state lines.[128]

---

deemed a branch and thus subjected to the strictures of the McFadden Act. In this connection, it should be noted that the Riegle-Neal Act contains an explicit prohibition against deposit production offices: a bank may not establish an out-of-state branch "primarily for the purpose of deposit production." Riegle-Neal Act § 109(a). *See infra* note 404 and accompanying text.

[123] *See* Felsenfeld, *supra* note 42, at 1050.

[124] Telephone banking transactions have increased 40% over the last five years, although full-fledged home banking (i.e., banking via personal computer and modem) remains experimental. Gibson, *supra* note 101, at 155.

[125] The term "deposit broker" is defined as "any person engaged in the business of placing deposits, or facilitating the placement of deposits, of third parties with insured depository institutions or the business of placing deposits with insured depository institutions for the purpose of selling interests in those deposits to third parties." 12 U.S.C. § 1831f(g)(1) (1994). *See* FAIC Securities, Inc. v. United States, 768 F.2d 352 (D.C. Cir. 1985) (invalidating regulation aimed at restricting deposit brokering).

[126] Michael Klausner, *An Economic Analysis of Bank Regulatory Reform*, 69 WASH. U. L.Q. 695, 728 (1991).

[127] "Any national banking association . . . may change the location of its main office to any authorized branch location within the limits of the city, town, or village in which it is situated, or, with a vote of shareholders owning two-thirds of the stock of such association for a relocation outside such limits and upon receipt of a certificate of approval from the Comptroller of the Currency, to any other location within or outside the limits of the city, town, or village in which it is located, but not more than thirty miles beyond such limits." 12 U.S.C. § 30(b) (1994).

[128] *See* Synovus Financial Corp. v. Board of Governors, 952 F.2d 426 (D.C. Cir. 1992)

Accordingly, in order to extend its network of branches into an adjacent state, a bank need only move its main office across the state line and then merge with an affiliated bank in the new state. Under the rules for branch retention following a merger or consolidation, the resulting bank may retain its branches in both states.[129] The OCC approved this practice in the 1994 case of First Fidelity Bank, N.A., Pennsylvania, which moved a few miles across the state line into New Jersey, merged with First Fidelity Bank, N.A., New Jersey, and then proceeded to operate branches in both states.[130] The Clinton administration supports exploitation of this loophole as a means of achieving interstate branching before the branching provision of the Riegle-Neal Act takes effect in 1997.[131] Thus, in the limited circumstances covered by 12 U.S.C. § 30(b), interstate branching is already the law despite the McFadden Act's apparent blanket prohibition on out-of-state branching.

H. *Summary*

To summarize, in many instances where the law has proven too rigid or simply outdated, the market has found ways to conduct banking operations across state lines. By employing a variety of devices—BHCs, correspondent relationships, ATMs, LPOs, deposit brokering—banks have succeeded in serving their customers' interstate needs. During debate over the Riegle-Neal Act, many commentators and even policymakers expressed the view that the question of interstate banking and branching had already been answered, not by Congress or the regulatory agencies, but by the market. "The battle was obviously over," remarked one observer, "so that Congress and the Clinton administration may be viewed as simply cleaning up some federal statutory debris that still clutters up the battlefield."[132] Although there remained geographic restrictions—enough to justify a thorough congressional review of the policy arguments for lifting

---

(permitting bank to move its main office less than 10 miles across state line from Alabama into Georgia, while retaining branch offices in Alabama).

[129] 12 U.S.C. § 36(b)(2) (1994).

[130] OCC Corporate Decision 94-04, 1994 OCC Ltr. LEXIS 4, Jan. 10, 1994.

[131] Arthur D. Postal, *Bankers Blink as OCC Tries to Give Go Ahead*, MORTGAGE MARKETPLACE, Jan. 17, 1994, at 1.

[132] Carter H. Golembe, *History Offers Some Clues on Significance of Interstate Branching*, BANKING POL'Y REP., Aug. 15, 1994, at 4.

them—the banks' circumvention tactics had already largely created a nationwide banking system.

III. The Deregulation Debate

In 1993 Congress turned its attention to the profound structural problem that it had declined to address in FDICIA, its last major piece of banking legislation. Bills proposing to allow interstate banking and branching by national banks were introduced in both the Senate and the House.[133] The bills were a frank admission that America had changed since the enactment of the McFadden Act in 1927 and the Douglas Amendment to the BHCA in 1956. Geographic limitations enacted when the country's banking needs and practices were relatively localized and unsophisticated were now difficult to defend. Congressman Neal underscored the point, reiterated by many lawmakers during the development of the new legislation, that the regulatory system needed to be modernized to reflect the changed circumstances:

> [T]he present Federal geographic constraints on banking were enacted in the 1930s and 1950s. The world has changed dramatically since then, and so has the way we do our banking in America. Americans are more mobile today than ever before. Technology now permits consumers to withdraw money from their accounts at ATMs across the country and around the world. Capital now flows from community to community and State to State irrespective of political boundaries.[134]

Indeed, the House and Senate bills were an acknowledgment of reality. Congress recognized that nationwide banking had come. It was clear that the appropriate legislative response to the market's initiatives was to eliminate the archaic laws that had long hindered the advance of the "global" banking model. It was time to do what had been urged in 1981, 1985, and 1991.[135] The Clinton administration supported deregulation, as had the Bush and Reagan administrations before it, and the bills met with bipartisan support in Congress.[136] The community banks, how-

---

[133] Interstate Banking and Branching Act of 1993, S. 371, 103d Cong., 1st Sess. (1993); Interstate Banking Efficiency Act of 1993, H.R. 2235, 103d Cong., 1st Sess. (1993).

[134] 140 Cong. Rec. H1851, 1858 (daily ed. Mar. 22, 1994) (statement of Representative Neal (D-Mass.).

[135] *See supra* notes 59–73 and accompanying text.

[136] Some lawmakers lamented this bipartisanship, on the ground that it prevented the

ever, together with those who feared that unbridled geographic expansion on the part of a few of the nation's most powerful banks would result in overconcentration,[137] a credit crunch for small businesses, and the draining of funds out of underprivileged communities, met the reformers' arguments with compelling counterarguments in support of the traditional barriers to bank expansion. Those arguments and counterarguments will now be considered.[138]

A. *Competition in the Financial Services Industry*

No other American industry is subject to geographic constraints as strict as those governing the banking industry. Coast-to-coast provision of goods and services is the norm in virtually every industry, and great value is assigned to nationwide standardization and consistency. Most enterprises are permitted to enter any market they wish. As one Congressman remarked, it is difficult to understand why banks should be treated differently: "Interstate commercial activities have long been the accepted mode of operation in this country. It has always seemed anomalous to me that products of every description could move so readily across State borders as a natural part of interstate commerce, but banking services could not."[139] Nor has concentration caused much alarm in other industries. Whereas in the banking industry a triumvirate of, say, Citibank, BankAmerica, and Wells Fargo would be viewed with horror, in other industries a small number of firms are allowed to dominate the market. As Congressman Neal noted, the same country that has 14,000 banks has only three automakers and three cereal manufacturers.[140] Moreover, the geographic limits on banking have serious efficiency

---

bill from making headlines. *See, e.g.*, 140 CONG. REC. H6774, 6779 (daily ed. Aug. 4, 1994) (statement of Representative Hoyer) ("Unfortunately, the broad consensus that exists for this bill may make it somewhat uninteresting to the media, where conflict is much preferred."). Indeed, the public seemed largely unaware of the debate over interstate banking, and the Riegle-Neal Act was passed without much public discourse.

[137] The term "concentration" reflects the amount of business conducted by the largest banks in a given market. In a highly concentrated market, a large percentage of the deposit and loan activity is carried on by a small number of institutions.

[138] Although the debate is now closed at the federal level, it continues in state legislatures on account of the opt-out provision. In this respect, the arguments reviewed here are still very much alive.

[139] 140 CONG. REC. H6782 (daily ed. Aug. 4, 1994) (statement of Representative LaFalce (D-N.Y.).

[140] *July 1993 Hearing*, *supra* note 120, at 32 (statement of Representative Neal).

consequences. In other industries, firms have benefited greatly from consolidation, cost-cutting, and economies of scale. "Would it be a good idea," asked Congressman Neal rhetorically, "to require that every Sears store in every city in this country [have] a separate board of directors?"[141] The BHC format, of course, requires precisely this kind of useless duplication of legal forms.[142] Whereas in other industries companies are free to expand to their optimal size while remaining single legal entities, banks are not.

The disparate treatment reserved for banks is particularly troubling in light of the fact that their nonbank competitors in the financial services industry are not subject to comparable geographic constraints. Although during the nineteenth century banks occupied a virtually unchallenged position as the principal providers of financial intermediation, today they face competition from a vast array of other enterprises, including insurance companies, investment companies, securities firms, investment banks, savings and loans (S&Ls), and finance companies.[143] Bank-like services are offered not just locally, but nationwide by financial conglomerates such as American Express, Merrill Lynch, and Sears Roebuck, as well as by industrial and retail companies such as General Motors, Ford, General Electric, J.C. Penney, and ITT.[144] Consumer savings are increasingly being invested in mutual funds instead of bank deposits, and large companies are turning with more frequency to the commercial paper markets[145] in order to satisfy their short-term capital needs.

To the extent that nonbanks offer mere extensions of credit, they cannot be viewed as usurping the banking function. But many nonbank products bear a disturbingly close resemblance to bank accounts. The cash management account (CMA) offered by Merrill Lynch, for example, which consists of a securities brokerage account, a money market fund, a checking account, and a credit card, is marketed as a high-interest alternative to the conventional bank account.[146] And, of course, Merrill Lynch has

---

[141] *Id.* (statement of Representative Neal).

[142] *See infra* note 152 and accompanying text.

[143] GAO 1993, *supra* note 10, at 27–29.

[144] Cohen 1985, *supra* note 38, at 975.

[145] The term "commercial paper" refers to short-term, unsecured promissory notes.

[146] Nestor, *supra* note 42, at 621. CMAs are generally considered not to constitute banking. *See, e.g.*, 1981 Op. Att'y Gen. Ore. 273 (Feb. 11, 1981).

the advantage of being free from geographic limitations. As a banking lawyer told the Senate Banking Committee in 1991:

> Merrill Lynch takes the equivalent of deposits and offers checking accounts at hundreds of offices across the country; Household International makes personal loans; Commercial Credit makes business loans; General Electric Credit makes more different kinds of loans than most banks; Western Union transfers money; AT&T offers credit cards; They all operate on an interstate basis but free of the crazy-quilt of geographic restrictions applicable only to full service commercial banks.[147]

As this testimony implies, banks cannot compete effectively against nonbanks when hemmed in by the McFadden Act and the Douglas Amendment, as well as the Glass-Steagall Act. "These outmoded Federal laws tilt the playing field against banks," noted Senator Dodd during floor debates over the Riegle-Neal Act.[148] Lifting the geographic restrictions imposed upon banks would level the playing field in the financial services industry. Those who fear that deregulation would result in overconcentration should bear in mind that, given the large number of nonbank financial organizations, the financial services industry would likely remain sufficiently diffuse.[149]

B. *Efficiency Concerns*

Proponents of deregulation argue that it would yield many benefits. First, and most broadly, there is a certain philosophical appeal to removing government-imposed barriers of any kind, to the extent that this can be accomplished without serious adverse consequences. Enabling bankers to decide the extent of bank expansion for themselves would be a victory for champions of the free market.[150] One congresswoman voiced support for the Riegle-Neal Act on such ideological grounds: "I believe that

---

[147] *Strengthening the Supervision and Regulation of the Depository Institutions: Hearings before the House Comm. on Banking, Housing, and Urban Affairs*, 102d Cong., 1st Sess. 446 (1991) [hereinafter *Mar. 1991 Hearing*] (testimony of Robert Carswell, partner, Shearman & Sterling).

[148] 139 CONG. REC. S1613 (daily ed. Feb. 16, 1993) (statement of Senator Dodd (D-Conn.)).

[149] Cohen 1985, *supra* note 38, at 976; Miller, *supra* note 32, at 1115. For further discussion of the debate concerning overconcentration, see *infra* notes 283–318 and accompanying text.

[150] *See* Carstensen, *supra* note 90, at 1414.

passage of this legislation is wholly consistent with the efforts of this Congress to reinvent government by eliminating outdated and unnecessary regulatory obstacles to economic growth."[151]

Deregulation may also produce substantial cost savings. In the absence of the McFadden Act, banks seeking to cross state lines would have the option of eschewing the BHC format and simply expanding via branching. The fuss and formalities of the BHC structure would thus be avoided: there would be no need for separately incorporated subsidiary banks with separate boards of directors, separate senior management, separate accounting ledgers, separate audits, separate regulatory reports and exams, separate capital requirements, and separate information systems.[152] Such duplication is expensive and inefficient. Consolidation of BHC subsidiaries into a single branch network would eliminate these areas of needless overlap. It would also facilitate the transfer of capital within a single family of banks, since § 23A of the Federal Reserve Act applies only to affiliated banks, not to branches.[153] In addition, unlike a family of affiliated banks, a single bank with multiple branches may engage in internal check clearing. When both payor and payee maintain accounts with the same bank, the bank can simply transfer the funds internally without using a clearinghouse.[154] All of these opportunities for reducing operating expenses make the prospect of interstate branching extremely attractive to banks.

NCNB (now NationsBank) Chairman Hugh McColl, an outspoken supporter of deregulation, told the Senate Banking Committee in 1991: "We project NCNB could save at least $20 million each year through interstate branching consolidations. Other multistate banking companies have estimated even higher annual cost savings."[155] By 1994 McColl had revised his estimate upward to $50 million per year.[156] Analysts predict that McColl's competitor in the West, BankAmerica Corp., would save $75 mil-

---

[151] 140 CONG. REC. H6780 (daily ed. Aug. 4, 1994) (statement of Representative Maloney (D-N.Y.)).

[152] Jane C. Linder & Dwight B. Crane, Bank Mergers: Integration and Profitability (1992) (unpublished paper), *in July 1993 Hearing*, *supra* note 120, at 228, 231; GAO 1993, *supra* note 10, at 92–99; Philip C. Meyer, *Branching Advocates Expect Significant Bottom Line Impact*, BANKING POL'Y REP., Aug. 15, 1994, at 20.

[153] GAO 1993, *supra* note 10, at 99.

[154] *See id.* at 67.

[155] *Mar. 1991 Hearing*, *supra* note 147, at 514 (testimony of Hugh McColl, chairman, NCNB Corp.).

[156] Jack Scis, *NationsBank Chief Responsible for Bill*, NEWS & RECORD, Nov. 14, 1994, at 7 *available in* LEXIS, News library, Curnws file.

lion in annual operating costs.[157] Indeed, a majority of commentators argue that a nationwide banking system could be developed at less cost if branching replaced the BHC as the primary means of crossing state lines.[158]

A few dissenters, however, maintain that the BHC structure is not substantially more expensive than branching. Some point out that the BHC format offers many of the same opportunities to pool resources that a branch network does.[159] The practice at BancOne and NationsBank, for example, is for the parent corporation to manage a standardized, centralized computer system for the benefit of all of its subsidiary banks.[160] Similarly, with regard to corporate strategy, the General Accounting Office (GAO) reports that "most subsidiary banks are already centrally managed; their primary policies are determined by the BHC . . . . It is unlikely that the consolidation of bank subsidiaries into branches would result in significant policy changes in the banking company."[161]

If the BHC structure is already serving some of the same ends that branching would serve, it is difficult to understand why there is a need to reform the law. It is doubtful that the BHC format is as inefficient as the reformers argue, given that in states allowing statewide branching, many banking organizations still prefer to use the BHC format in-state rather than fusing the subsidiaries into a single bank.[162] The BHC strategy clearly has its advantages. Each bank subsidiary, operating at most on a statewide basis, is in effect a local bank, with a board of directors and senior managers drawn from the local community and sensitive to that community's needs.[163] Consumers often prefer dealing with local businesses, and may particularly distrust financial institutions managed by outsiders.[164] The highly successful BancOne Corp., an Ohio-based BHC with banks in fourteen states, favors the BHC format for precisely this reason. As a BancOne executive explained: "Our feeling is that an independent

---

[157] *See* Thomas Hoffman, *Capitalizing on Client/Server*, COMPUTERWORLD, Feb. 1, 1995, at 53.

[158] *See, e.g.*, Byrne & Coulter, *supra* note 72, at 681; Cohen, *supra* note 79, at 9; Miller, *supra* note 32, at 1101–02.

[159] *See, e.g.*, GAO 1993, *supra* note 10, at 48–49.

[160] Wilmarth, *supra* note 9, at 1006.

[161] GAO 1993, *supra* note 10, at 48–49.

[162] Wilmarth, *supra* note 9, at 1007; Savage, *supra* note 51, at 1081.

[163] Savage, *supra* note 51, at 1081.

[164] Wilmarth, *supra* note 9, at 1043.

bank with an independent president serving the community creates value which more than offsets the increased expenses of maintaining separate entities."[165] In sum, although branches are concededly less expensive than bank subsidiaries, there is disagreement about how great the difference is and how much it matters.

Branching would, however, afford more opportunities for cost rationalization, and the efficiency gains would be especially great if, as has been widely forecast and indeed has begun to occur,[166] relaxation of the geographic restrictions resulted in increased merger and acquisition activity. Mergers are typically accompanied by corporate streamlining and back-office consolidation. As of 1994, estimates of the cost savings associated with bank acquisitions were in the fifteen to thirty percent range, with some commentators asserting an average of thirty-five percent for in-market mergers. Even fifty percent was not considered unusual.[167] The mergers of the past few years, such as BankAmerica-Security Pacific, Chemical-Manufacturers Hanover, Society-Ameritrust, and First Union-Southeast, have all produced substantial reductions in overhead expenses.[168] For example, after their 1991 merger, Chemical and Manufacturers Hanover realized considerable savings by closing a large number of check-clearing centers.[169] To be sure, the savings associated with mergers are probably more the result of elimination of waste than exploitation of economies of scale. Sheer size is not the key to profitable banking.[170] The

---

[165] Steve Klinkerman, *Wider Branching Welcomed But Seen as No Quick Fix*, AM. BANKER, Mar. 8, 1991, at 1, *available in* LEXIS, Bankng Library, Ambank File, (quoting William Boardman, executive vice president, BancOne Corp.). *See also* Stephen A. Rhoades & Donald T. Savage, *Interstate Branching: A Cost-Saving Alternative?*, BANKERS MAG., July-Aug. 1993, at 35.

[166] *See, e.g.*, GAO 1993, *supra* note 10, at 47–48. For discussions of potential cost savings from recent merger activities, see, e.g., James R. Kraus, *Savings, Technology, and Clout Drove Chase-Chemical Merger*, AM. BANKER, Aug. 29, 1995, at 1, *available in* LEXIS, Bankng Library, Ambank File; Christine Dugas & David Henry, *Big Year for Bank Deals: First Interstate Faces $10 Billion Unsolicited Bid*, USA TODAY, Oct. 19, 1995, at B1.

[167] Marilyn R. Seyman & Steven P. Williams, *Comment: To Stay Independent, Check Cost Structures*, AM. BANKER, Jan. 6, 1995, at 9, *available in* LEXIS, Bankng Library, Ambank File; MCCOY ET AL., *supra* note 51, at 128–29.

[168] MCCOY ET AL., *supra* note 51, at 52.

[169] Gibson, *supra* note 101, at 162.

[170] *See* JOHN H. BOYD & STANLEY L. GRAHAM, INVESTIGATING THE BANKING CONSOLIDATION TREND (1991) *reprinted in July 1993 Hearing*, *supra* note 120, at 184. There has always been a myth, however, that size is critical. The banking industry has historically emphasized size over profitability. *See* MCCOY ET AL., *supra* note 51, at 10. Consumers, too, seem to favor big banks, primarily for safety and soundness reasons. The popular perception is that the bigger the bank, and the better known its

general consensus is that there are increasing returns to scale as banks approach $100 million in assets, but none thereafter. In fact, extremely large institutions (i.e., those having assets of $10 billion or more) commonly experience diseconomies of scale.[171] Nevertheless, many prominent commentators have taken the side of Hugh McColl and his reform-minded colleagues. Banking consultants have offered optimistic predictions about the efficiency gains that would be realized if industry consolidation were accelerated through liberalization of the branching rules. Lowell Bryan of McKinsey & Co. believes that nationwide branching could save the industry $10 billion per year in operating costs,[172] and others have proposed figures as high as $23 billion.[173] Given the consultants' rosy forecasts, as well as the optimism expressed by many in the banking industry itself, it seems clear that interstate branching should at least be made an option, so that those banking organizations that are confident they could cut costs would have the opportunity to experiment with the branching format.[174]

Cost-efficiency is also critical to the safety and soundness of the banking system. Recognizing that this was a selling point of the interstate branching bill, McColl told the Senate Banking Committee:

> These significant savings will go directly to the bottom lines of our nations' banks. Higher retained earnings will strengthen capital ratios. Stronger capital ratios will rebuild confidence in the banking system. And, most important, a well-capitalized system will give my colleagues the assurance they need to make even more loans to creditworthy borrowers.[175]

Regulators have made the same point. In addition to simplifying regulatory oversight of the banking industry—since there would

---

name, the less likely it is to fail. Alternatively, the bigger the bank, the more likely the government is to view it as "too big to fail." *See* Wilmarth, *supra* note 9, at 1022, 1043.

[171] Litan, *supra* note 55, at 529; Wilmarth, *supra* note 9, at 1007–08.

[172] Lowell Bryan, *A Blueprint for Financial Reconstruction*, HARV. BUS. REV. May-June 1991, at 81.

[173] This estimate is attributed to consultant Sanford Rose. *See* Litan, *supra* note 55, at 530.

[174] Many commentators emphasize the importance of experimentation. *See, e.g.*, Miller, *supra* note 32, at 1101 ("[I]t would be better to allow the market actually to experiment with additional geographic bank expansion, including interstate branching. If interstate branching proves to be efficient and profitable, it will survive and flourish . . ."); GAO 1993, *supra* note 10, at 80.

[175] *Mar. 1991 Hearing*, *supra* note 147, at 514 (testimony of Hugh McColl, chairman, NCNB Corp.).

be only one quarterly report per banking company instead of dozens (one for each separate bank subsidiary)[176]—interstate branching would enhance the stability of the banking system. The Comptroller of the Currency, Eugene Ludwig, testified to this effect: "To the extent that banks realize cost savings, they will be able to augment capital, directly strengthening the bottom line of both individual institutions and the industry as a whole, and thus directly improving safety and soundness."[177] In sum, the barriers to geographic expansion should be lifted in the interest of bank stability, and for the good of the economy as a whole. The present practice of conducting interstate banking using the roundabout methods discussed in Part II, rather than the direct method of branching, imposes substantial costs on the public—costs that many argue are not worth incurring.[178]

C. *Consumer Convenience and Credit Availability*

The point of helping banks reduce their operating expenses is not, of course, to make bankers rich, but to benefit consumers and the economy as a whole. The hope is that banks will pass their cost savings along to consumers in the form of lower prices and greater credit availability, as they can be expected to do if the market remains competitive.[179] Senator Dodd advertised his 1993 proposal on exactly these terms: "The interstate branching bill I am introducing today will enable banks to eliminate . . . duplication and waste, and every dollar saved is a dollar that can be used to make loans and reduce the credit squeeze."[180] The

---

[176] *See* Letter from Robert D. Reischauer, director, Congressional Budget Office, to Representative Henry B. Gonzalez (D-Tex.) (Mar. 22, 1994), *in* H.R. REP. No. 103-448, 103d Cong., 2d Sess. 31 (1994); GAO 1993, *supra* note 10, at 78–79, 93. *But see The Effect of Interstate Branching on National, State and Local Economies, Hearing Before the Subcomm. on Economic Stabilization, Comm. on Banking, Finance and Urban Affairs*, 102d Cong., 1st Sess. 27 (1991) [hereinafter *May 1991 Hearing*] (testimony of Keith Ellis, Bank Commissioner, Delaware) ("While a majority of State and Federal agencies have entered information-sharing agreements, information available to regulators will be reduced in an interstate branching environment as the conversion of affiliates to branches reduces the number of audits and examinations. This is one of the cost savings which proponents of interstate branching hope to effect, although reduced oversight in the name of efficiency may not be a desirable public policy goal.").

[177] *Interstate Banking and Branching: Hearing Before the Subcomm. on Financial Institutions Supervision, Regulation and Deposit Insurance of the House Comm. on Banking, Finance and Urban Affairs*, 103d Cong., 1st Sess. 16 (1993) [hereinafter *Oct. 1993 Hearing*] (testimony of Eugene A. Ludwig, Comptroller of the Currency).

[178] Felsenfeld, *supra* note 42, at 1059.

[179] Litan, *supra* note 55, at 532, 542; GAO 1993, *supra* note 10, at 80, 103.

[180] 139 CONG. REC. S1613 (daily ed. Feb. 16, 1993) (statement of Senator Dodd).

theme of freeing up funds for lending was repeatedly sounded during congressional debates on the Riegle-Neal legislation. Congressman Hoyer, for example, sketched a concrete picture of how deregulation would help consumers and small businesses: "Every dollar the bank ties up in regulatory capital requirements, or pays in administrative costs, is a dollar that cannot be loaned to a small business which wants to expand its operations; or to a family which is paying for a child's college education."[181] Interstate branching, in other words, is being touted as a solution to the credit crunch. This potential increase in the availability of capital to fund innovative projects and ideas, particularly those of small business, could be vital to the nation's economic health.[182]

Big business, for its part, would also benefit from interstate branching. Instead of maintaining dozens or even hundreds of deposit accounts with different banks across the country, large corporations could consolidate accounts and thereby reduce monitoring costs, transfer fees, and other expenses.[183] An executive of Occidental Petroleum, for example, testified before a House subcommittee that under an interstate branching regime his company would save three to four percent per year in banking costs.[184] Large companies would also benefit as borrowers, because the bigger the bank, the higher its limit on loans to a single borrower.[185] Occidental predicted significant reductions in transaction costs if it were permitted to borrow larger amounts from a smaller number of banks.[186]

---

[181] 140 CONG. REC. H1860 (daily ed. Mar. 22, 1994) (statement of Representative Hoyer).

[182] *See* LARRY A. FRIEDER, BANK LEGISLATION IN A RAPIDLY CHANGING ENVIRONMENT: IS THE STATUS QUO ADVISABLE? *reprinted in Interstate Banking and Branching: Hearings Before the Subcomm. on Financial Institutions, Supervision, Regulation and Deposit Insurance of the House Comm. on Banking, Finance and Urban Affairs*, 103d Cong. 1st Sess. 43 (1993) [hereinafter *Sept. 1993 Hearing*], at 157 ("[M]ost regional and superregional banks have increased their strategic interest in small business lending. This new and large focus should bode very well for the country given the production and job generating abilities of this sector.").

[183] *See* Neal & McCollum, *supra* note 105, at 59–60; *July 1993 Hearing*, *supra* note 120, at 6 (testimony of Robert J. Pisapia, Occidental Petroleum Corp.) ("For example, if OXY were to have a single creditor bank with branches in Chicago, Pittsburgh, Dallas, Charlotte, and Houston we could be using 1 bank, instead of 10 and a few bank accounts instead of many.").

[184] *July 1993 Hearing*, *supra* note 120, at 2 (testimony of Robert J. Pisapia).

[185] A national bank may not make an unsecured loan to a single borrower in an amount exceeding 15% of its unimpaired capital and unimpaired surplus. The bank may lend an additional 10% of its unimpaired capital and unimpaired surplus to that same borrower, provided the loan is fully secured. 12 U.S.C. § 84(a) (1994). *See* Litan, *supra* note 55, at 540.

[186] *July 1993 Hearing*, *supra* note 120, at 6 (testimony of Robert J. Pisapia)

Interstate branching would provide consumers with not only lower-cost services but also greatly enhanced convenience. The balkanized, fragmented character of the current American banking system makes it impossible to obtain services and execute transactions that in other countries would be considered basic. Whereas a Canadian can walk into a full-service branch of her own bank in any part of the country, an American finds herself with no banking support as soon as she strays from home. With the exception of cash withdrawal, she can accomplish virtually nothing outside her home state (or, in the thirteen states restricting branching, outside her county or other political subdivision). She cannot cash or deposit a check, seek banking advice, or open or close an account.[187] If she moves, she has to find a new bank, and if her children go to college out-of-state, they have to open their own local accounts.[188] Frequent relocations are increasingly common. One commentator told a House subcommittee: "As much as 30 percent of a bank's customers turn over every year, so there is mobility. To have banks that are branched and ready to accommodate these changes has a very positive impact."[189]

The branching restrictions also cause serious inconvenience to business travelers and tourists, who comprise a large percentage of the increasingly mobile American population,[190] as well as to people who split their time between two or more states. Many consumers fall into the latter category quite by chance, simply because they happen to live in a multistate metropolitan area such as Washington, D.C., Philadelphia, St. Louis, or greater New York City (often called the "Tri-State Area").[191] The line-drawing mandated by the McFadden Act seems particularly arbitrary—even absurd—as applied to such urban centers, where crossing state lines is part of the daily routine. Four million Americans commute across state lines every day, and sixty million (i.e., nearly one-quarter of the population) live in multistate metropolitan areas.[192] Interstate branching would solve the prob-

("Occidental bucks up against some of those legal lending limits . . . . [I]f the banks are able to increase their limit to us we would be able to condense our network because operating business would be spread amongst fewer creditor banks.").

[187] Miller, *supra* note 32, at 1108–09.

[188] GAO 1993, *supra* note 10, at 102.

[189] *See Sept. 1993 Hearing*, *supra* note 182, at 48 (testimony of Larry A. Frieder, professor, Florida A & M University).

[190] Neal & McCollum, *supra* note 105, at 57.

[191] *See* GAO 1993, *supra* note 10, at 103.

[192] 140 CONG. REC. H1858 (daily ed. Mar. 22, 1994) (statement of Representative Neal); GAO 1993, *supra* note 10, at 89.

lem by allowing urban banks to follow customers into suburbs located in neighboring states.[193] It would also facilitate interstate commerce and help companies that operate in multiple states. As McColl noted, "In terms of convenience, interstate branching will allow a Houston-based company doing business with NCNB to issue paychecks to employees in other states that can be deposited and given immediate credit at NCNB, illustratively, in Columbia, South Carolina or Orlando, Florida."[194] Thus, the reformers make a convincing case that American consumers continue to experience considerable inconvenience as a result of branching restrictions and would benefit greatly from an interstate branching regime.

D. *Diversification of Bank Assets and Liabilities*

A further argument in favor of interstate branching—one of the most compelling and frequently advanced—concerns diversification of bank assets and liabilities. A bank that conducts all of its lending and deposit-collection activity within a single, narrowly defined region inevitably becomes dependent on that region's economic health. In periods of crisis, the bank has nowhere else to turn and must endure whatever hardships the local economy faces. Dependence on a single region poses threats to both sides of a bank's balance sheet. The bank is exposed to both liquidity risk (i.e., the risk that its deposit base will dry up, leaving it with no funds to lend, or that many depositors will demand their money at once, causing a run) and credit risk (i.e., the risk that many borrowers, all crippled by the same regional recession, will default on their loans simultaneously).[195]

Liquidity risk can be controlled by attracting deposits not only from businesses but also from individual consumers, who are less likely to make withdrawals based on small fluctuations in interest rates.[196] Congressmen Neal and McCollum emphasized this point in their 1994 report:

> [B]ranch banking enables banks to cast a wider net for core deposits, the name often given to the basic transaction ac-

---

[193] *See* Savage, *supra* note 51, at 28.

[194] *Mar. 1991 Hearing*, *supra* note 147, at 515 (testimony of Hugh McColl, chairman, NCNB Corp.).

[195] GAO 1993, *supra* note 10, at 74–77; Nellie Liang & Stephen A. Rhoades, *Geographic Diversification and Risk in Banking*, 40 J. ECON. & BUS. 271, 273 (1988).

[196] Liang & Rhoades, *supra* note 195, at 275; Wilmarth, *supra* note 9, at 984.

> counts of consumers and businesses alike . . . . Branching allows a bank to be accessible to more of both types of customers, and thereby collect more deposits, and core deposits are cheaper and more stable sources of funds than are CDs or brokered funds.[197]

Similarly, on the asset side, the McFadden Act effectively forces banks to violate one of the fundamental principles of modern economics—portfolio theory. According to portfolio theory, investments should be spread across a broad array of enterprises, industries, and geographic regions.[198] Since different classes of investments should rise and fall in value independently of one another, gains in one area should offset losses in another, resulting in a stable portfolio. By contrast, an undiversified portfolio may become worthless in an instant on account of the covariance principle (i.e., the tendency of investments of a particular type to rise and fall together).[199] A bank's investment portfolio consists of its loans. Consistent with portfolio theory, if a bank is to minimize its risk of insolvency, those loans should be made to enterprises from different industries and different parts of the country.[200] Empirical evidence supports this hypothesis. In a study of more than 5500 banking organizations over a ten-year period, economists with the Federal Reserve determined that bank risk is substantially reduced through diversification of loan portfolios.[201]

A few examples will suffice to demonstrate, however, that the law has often hampered such diversification, with grave consequences flowing from banks' overdependence on local economies. States restricting in-state branching have typically experienced the highest rates of bank failure.[202] Illinois, for example, long prohibited banks from maintaining branches of any kind. Pursuant to the Illinois unit-banking statute, a bank could operate no more than a single office.[203] Limited to a single storefront,

---

[197] Neal & McCollum, *supra* note 105, at 60.

[198] *See* RICHARD A. BREALEY & STEWART C. MYERS, PRINCIPLES OF CORPORATE FINANCE 155–74 (1991).

[199] Freeman, *supra* note 96, at 1143.

[200] *See* Wilmarth, *supra* note 9, at 983–84.

[201] Liang & Rhoades, *supra* note 195, at 278. This study was based on geographic expansion in-state, not across state lines. Liang and Rhoades expect that their results would be even more dramatic if the data were based on interstate expansion. *Id.* at 271–72.

[202] McCOY ET AL., *supra* note 51, at 63; Miller, *supra* note 32, at 1104.

[203] "No bank shall establish or maintain more than one banking house, or receive deposits or pay checks at any other place than such banking house, and no bank shall establish or maintain in this or any other state of the United States any branch bank,

banks such as Continental Illinois were unable to build a broad retail deposit base. Continental's deposits came chiefly from inside the narrow Chicago Loop. Beyond the Loop, its funds primarily came from large-denomination, uninsured CDs held by a small number of big institutional depositors, some of them foreign.[204] This was a risky way to organize the liabilities side of its balance sheet. On the assets side, Continental pursued an equally perilous strategy. Its chief attempt at diversification was to acquire more than one billion dollars in loans from the defunct Penn Square Bank of Oklahoma (whose failure was also due in part to state branching restrictions).[205] These assets, however, were not only low-quality but were also confined to a single industry—the unstable oil and gas industry.[206] When Continental announced major loan write-offs in 1984 and appeared ready to collapse, the institutional depositors on which it had become dependent panicked and triggered a run. In the end, Continental had to be rescued by the federal government in one of the largest bailouts in American history.[207] Commentators often attribute this debacle to Illinois's straitjacket unit-banking statute,[208] which has since been amended to permit limited branching.[209]

---

nor shall it establish or maintain in this State any branch office or additional office or agency for the purpose of conducting any of its business." Illinois Banking Act § 6, Ill. Ann. Stat. ch. 17, § 313 (Smith-Hurd 1981).

[204] Cohen 1985, *supra* note 38, at 985; Laurie S. Goodman & Sherrill Shaffer, *The Economics of Deposit Insurance: A Critical Evaluation of Proposed Reforms*, 2 YALE J. ON REG. 145, 151 n.33 (1984).

[205] *See* David LaGesse, *Poole's Pocket Guide to Failure: Ingredients Are Mismanagement, Archaic Laws*, AM. BANKER, July 2, 1987, at 11, *available in* LEXIS, Bankng Library, Ambank File.

[206] Cohen 1985, *supra* note 38, at 985; Goodman & Shaffer, *supra* note 204, at 151 n.31.

[207] Goodman & Shaffer, *supra* note 204, at 151 n.31.

[208] *See, e.g.*, GAO 1993, *supra* note 10, at 76; Cohen 1985, *supra* note 38, at 984–85; Marian Courtney, *Lessened Risk Seen for Banks*, N.Y. TIMES, Mar. 23, 1986, at 11NJ6; *Freeing American Banks*, ECONOMIST, June 22, 1985, at 15; David S. Holland, *A Broad Banking Bill This Year? A Prediction*, BANKING EXPANSION REP., Sept. 3, 1984, at 1.

[209]

> A bank organized under this Act or subject hereto . . . shall, without specific mention thereof in the charter, have all the powers conferred by this Act and the following additional general corporate powers: . . . (15) To establish and maintain, in addition to the main banking premises, branches offering all banking services permitted at the main banking premises . . . (a) A bank organized under this Act or subject hereto may establish and maintain: (i) not more than 10 branches in the home county of the establishing and maintaining bank; (ii) not more than 5 branches in each contiguous county of the establishing and maintaining bank; and (iii) not more than 5 branches in Illinois that are not in the home county or in a contiguous county of the establihsing

A similar situation obtained in Texas in the mid-1980s. When nine of the ten largest banks in the state failed or had to be rescued,[210] lack of diversification was once again largely to blame. Texas had long been a unit-banking state, as well as a state deeply dependent on the energy industry. Most bank assets were tied to oil, or to Texas agriculture and real estate.[211] When the state's economy entered a sustained period of crisis, its banks were devastated. Several massive bailouts followed, including two of the costliest in American history (MCorp and First Republic Bank Corp.), and out-of-state banks moved in. Today, Texas has no big banks of its own; the market is dominated by outsiders, the top five of which (NationsBank, Chemical, BancOne, BankAmerica, and First Interstate) held more than one third of the state's deposits as of 1994.[212] Commentators and regulators have been quick to recognize the role that branching restrictions played in precipitating the Texas disaster.[213] Robert Clarke, then Comptroller of the Currency, put it plainly: "[T]he laws made it hard to get more diversified."[214] Once the damage had been done, Texas identified its diversification problem and took steps to rectify it. Banks were allowed to branch county-wide (by constitutional amendment in 1986)[215] and eventually statewide.[216]

---

and maintaining bank, but no such branch shall be located more than 10 miles from the main banking premises of the establishing and maintaining bank.

ILL. COMP. STAT. ANN. ch. 205, act 5, § 5 (Smith-Hurd 1995).

[210] James Bates, *Boston Bank Failure Eerily Similar to Texas Debacle*, L.A. TIMES, Jan. 8, 1991, at A1.

[211] Sarah Peterson, *Texas, Its Economy Standing Less Tall, Bows to Outside Banking*, CHRISTIAN SCI. MONITOR, Jan. 30, 1987, at 19.

[212] Arnold G. Danielson, *Impact of Nationwide Banking in the South May Not Be Dramatic*, BANKING POL'Y REP., Aug. 15, 1994, at 11. It is ironic that Texas banks were among those that the regional compacts of the mid-1980s were designed to exclude. As it turned out, Texas banks could barely sustain themselves, let alone conduct extensive acquisition campaigns. *See* MCCOY ET AL., *supra* note 51, at 60; GAO 1993, *supra* note 10, at 154–55.

[213] *See, e.g.*, Litan, *supra* note 55, at 533; *cf.* Wilmarth, *supra* note 9, at 983.

[214] *See* David LaGesse, *Despite Woes, Texans Call for More Deregulation*, AM. BANKER, Oct. 28, 1987, at 3, *available in* LEXIS, Bankng Library, Ambank File, (quoting Robert L. Clarke, Comptroller of the Currency).

[215] TEX. CONST., art. XVI, § 16.

[216] Following Texas v. Clarke, 690 F.Supp. 573 (W.D. Tex. 1988), in which a federal district court upheld an OCC decision to allow two national banks in Texas to branch outside their home counties; the Texas Banking Commission authorized state-chartered banks to do the same, despite the constitutional prohibition on branching across county lines. *See Texas Regulators Give State Banks Right to Branch Across County Lines*, BNA'S BANKING REP., Sept. 26, 1988, at 538.

Similarly, New England became "the Texas of 1990s"[217] when a regional economic downturn brought about the collapse of fifty-three banks and S&Ls in 1991.[218] Although several New England BHCs owned banks throughout the region—the Bank of New England Corp. (BNE), for example, operated a total of 320 branches through bank subsidiaries in Massachusetts, Connecticut, and Maine[219]—the loan portfolios of the individual banks were not sufficiently diversified to withstand the severe regional recession of the early 1990s. The most notorious failure, that of BNE, can be attributed in large measure to the bank's holding too high a concentration of local real estate loans.[220] Richard Syron, President of the Federal Reserve Bank of Boston, explained the problem this way: "The experience in New England demonstrates that if we had full interstate banking, institutions would have had diversified portfolios, thus diminishing the problems we have had over the last several years."[221]

Merely branching into contiguous states, however, would probably not have helped BNE, as the entire region was economically depressed in this period. In order to secure the full benefits of diversification, banks must extend their reach not just slightly, but substantially. Just as statewide branching is often insufficient, especially in one-crop or one-mineral states,[222] multistate branching also fails to permit adequate portfolio diversification when the states are neighbors or share the same principal industries. In order to minimize insolvency risk, banks should lend into distant regions and across a broad spectrum of industries.[223] Indeed, multiregional expansion appears to be the formula for success in the 1990s. In a 1994 book on banking strategy, three experts recommended:

> [P]erhaps the best way to ensure a truly diversified business mix is to do mergers between banks in regions whose core business activities are not highly correlated . . . . KeyCorp

---

[217] Peter G. Gosselin & Doug Bailey, *The Day Bank of New England Failed*, BOSTON GLOBE, Jan. 7, 1991, at 1.

[218] Arnold G. Danielson, *New England Banking: Great Opportunities for the Survivors*, BANKING POL'Y REP., Feb. 3, 1992, at 6.

[219] Stephen Labaton, *Regulators Pick Buyer to Operate New England Bank*, N.Y. TIMES, Apr. 23, 1991, at A1.

[220] Danielson, *supra* note 218, at 6.

[221] *June 1993 Hearing*, *supra* note 64, at 25 (testimony of Richard F. Syron, president, Federal Reserve Bank of Boston).

[222] Cohen 1985, *supra* note 38, at 987.

[223] MCCOY ET AL., *supra* note 51, at 130–31; GAO 1993, *supra* note 10, at 74–76.

> and Fleet Financial are examples of banks that have deliberately pursued acquisitions for the purpose of diversification. KeyCorp [of New York] has pursued its "polar banking" or "snowbelt" strategy by buying banks in Maine, Oregon, Washington, and Alaska . . . . Fleet Financial [of Rhode Island] has long been sensitive about its dependence on its New England roots. By merging with Norstar of upstate New York and developing national product-based businesses such as mortgage banking and student loans, it obtained some measure of diversification.[224]

Adoption of such a multiregional philosophy would not only improve the bottom line of individual banks but would also reduce the likelihood of the BIF's having to withstand another failure like BNE's.

The BNE episode also demonstrates that diversification at the BHC level is not sufficient. Each individual bank must be adequately diversified. Loans made by Bank *X*'s affiliate, Bank *Y*, do not help *X* diversify its portfolio. Since *X* and *Y* have separate balance sheets, their loan losses cannot be spread over the two combined. Only when good assets are included in *X*'s own portfolio will the income they produce offset losses generated by *X*'s bad assets.[225] When times are prosperous in *Y*'s region, *Y* can only assist *X* to the limited extent allowed by the affiliated transaction rules considered in Part II.[226]

Alternatively, *X* and *Y* might swap loans. But if they originated loans for one another on a regular basis, they would likely be deemed *de facto* branches, and the arrangement would thus violate the McFadden Act's prohibition against interstate branching.[227] Nor is the purchase of loan participations[228] the optimal way to achieve geographic diversification, because the acquiring bank does not know the borrower and has not made its own credit analysis.[229] It did not help Continental Illinois, for example, to purchase shaky energy loans in the South.[230] A Treasury Department official explained:

---

[224] McCoy ET AL., *supra* note 51, at 131.

[225] Freeman, *supra* note 96, at 1145.

[226] *See supra* note 87 and accompanying text.

[227] Freeman, *supra* note 96, at 1146. See *supra* note 83 and accompanying text for a discussion of *de facto* branches.

[228] *See supra* note 109 and accompanying text for a discussion of this technique for circumventing the McFadden Act.

[229] *Id.* at 1147. *But see* Carstensen, *supra* note 90, at 1407.

[230] Cohen 1985, *supra* note 38, at 985. *See supra* text accompanying notes 205–206.

> Lending profitably means knowing the opportunities, knowing the customers . . . . [T]he problem with some banks that were forced to buy loans because they didn't have offices where the loans were produced, they bought hot money loans that they didn't understand. The difference would be, if they had operations in those States, they could make the safe loans.[231]

Thus, in order to ensure maximum financial safety and soundness, a bank should originate loans through its own full-service branches in other states.

Strictly speaking, portfolio theory contemplates an individual investor purchasing small shares of many businesses, whereas a bank seeking to diversify its loans and deposits acquires entire companies (i.e., other banks). Thus, portfolio theory is not perfectly applicable to the latter situation.[232] The diversification gains produced by interstate branching will only be valuable to the extent they are not offset by increased costs resulting from expansion into new regions, acquisitions of other institutions (or establishment of a presence *de novo*),[233] and management of the resulting enterprise. Accordingly, Liang and Rhoades qualify the conclusion of their 5500-bank study by warning that "increased operating risk may accompany increased diversification and reduced financial risk."[234] First, it may prove difficult to assimilate target banks into the consolidated enterprise, as corporate cultures may clash. Second, customers, often wary of changes of ownership, may react by turning away.[235] Third, a newly expanded bank must exercise particular caution in making credit decisions at the outset, when it may not be as familiar with each of its markets as a small institution operating in only one market.[236] Finally, the bank is certain to experience substantial increases in management costs since far-flung enterprises are difficult to operate.[237] However, in light of advances in communications

---

[231] *May 1991 Hearing*, *supra* note 176, at 11 (testimony of Robert R. Glauber, Under Secretary (Finance), U.S. Dep't of the Treasury).

[232] Liang & Rhoades, *supra* note 195, at 272.

[233] A bank may expand into a new region either by acquiring an existing institution or by branching anew ("*de novo*"). The distinction is important for the purposes of the Riegle-Neal Act. *See infra* notes 349–353.

[234] *Id.* at 282.

[235] *See* Carstensen, *supra* note 90, at 1417; Linder & Crane, *supra* note 152, at 8, 18–19; Wilmarth, *supra* note 9, at 987.

[236] Cohen 1985, *supra* note 38, at 985; GAO 1993, *supra* note 10, at 114.

[237] *See* Carstensen, *supra* note 90, at 1415; Elizabeth S. Laderman & Randall J. Pozdena, *Interstate Banking and Competition: Evidence from the Behavior of Stock Returns*, ECON. REV., FEDERAL RESERVE BANK OF SAN FRANCISCO 32 (1991), at 34.

and transportation, it appears that a multistate bank can be managed soundly—especially given the recent successes of BHCs such as NationsBank and BancOne in Texas, as well as Norwest and First Bank System, Minnesota BHCs that own subsidiary banks in Colorado.[238] Thus, if a bank can successfully meet the challenges posed by interstate expansion, it can reap the substantial portfolio benefits of diversification.

E. *Community Commitment and Access to Credit*

There remains, however, the question whether giant financial institutions of the kind envisaged by the reformers are inherently bad. A minority of commentators and scholars argue vehemently that geographic restrictions serve an important purpose and that deregulation would damage both the banking system and the economy as a whole. Repeal of the McFadden Act and Douglas Amendment, they contend, would allow a few behemoths to rise quickly to power and dominate the banking industry. Under the new regime—a regime of big money, standardization, and disregard for local economies—the noble American traditions of decentralization and dispersed ownership of the means of financial intermediation would be lost. As discussed in Part I, the defenders of decentralization have often couched their arguments in moral terms.[239] The following view, articulated during the heyday of BHC expansion in the mid-1980s, is representative:

> We should be proud of our decentralized finance and capitalism. It helps promote efficiency, competitive pricing and interest rates, along with fair access to credit for small business, housing, and other needs. Decentralization fosters social mobility, political pluralism, economic development and self-respect all over the country.[240]

According to this view, deregulation would threaten deeply held American values.

Americans have long believed that banks should be accountable to their communities. One of the axioms of American banking regulation is that banks are responsible for making credit available to small businesses and other local borrowers even

---

[238] McCoy et al., *supra* note 51, at 131.

[239] *See supra* text accompanying notes 63–64.

[240] William A. Lovett, *Federalism, Boundary Conflicts and Responsible Financial Regulation*, 18 Loy. L.A. L. Rev. 1053, 1056 (1985).

when it may be more profitable for them to invest elsewhere. Under a nationwide branching regime, argue opponents of reform, banks would likely lose touch with their communities. Among other things, this would lead to a reduction of credit availability for small businesses, which are heavily dependent on bank loans since they cannot raise money by issuing securities in the public markets.[241] Small businesspeople often obtain loans on the strength of the personal bonds of trust that they have established with community bankers. If a sole proprietorship or other small enterprise cannot satisfy the standardized lending criteria and collateral requirements of a large, impersonal multistate bank,[242] it may be able to count on the "character lending" practices of a local institution (i.e., the extension of credit based largely on faith in the borrower's good character).[243] Unlike the branch manager of a massive out-of-state institution—who ultimately has little authority and, as one consumers' advocate put it, "may be transferred to Alaska tomorrow"[244]—an independent banker can rely on her instinct and intimate knowledge of the community in evaluating loan applications. She may also be more willing to make small loans. Indeed, it was for this reason that the president of a small-town bank in Pennsylvania advised Congress not to lift the geographic restrictions: "I make business loans for $5,000, $2,500, and those kinds of loans will go away. Big, large banks are not interested in those kinds of loans for those businesses."[245]

On the other hand, the GAO predicted in its 1993 study that small businesses would benefit from interstate branching, both because they would have greater choice in banking services and because large banks, whose major corporate customers are able to go directly to the capital markets to raise funds, are increasingly targeting small businesses as potential borrowers.[246] Simi-

---

[241] GAO 1994, *supra* note 58, at 15.

[242] *See id.* at 5 n.18 ("Under a centralized and standardized system, loan officers working in a central location make loans according to standardized financial criteria. This system may depersonalize the relationship between the loan officer and borrower, making it difficult for the loan officer to take into account relevant credit information that is not captured using standardized criteria.").

[243] *See* McCoy ET AL., *supra* note 51, at 250; GAO 1993, *supra* note 10, at 119; Wilmarth, *supra* note 9, at 1038–39.

[244] *July 1993 Hearing*, *supra* note 120, at 18 (testimony of Chris Lewis, Consumer Federation of America).

[245] *Id.* at 15 (testimony of James R. Lauffer, chairman, president, and CEO, First National Bank of Herminie).

[246] GAO 1993, *supra* note 10, at 118.

larly, in its 1994 report, the GAO noted that the efficiency gains associated with interstate expansion and consolidation would free up more funds for small business lending, and that centralization and standardization help make such lending more profitable.[247] Perhaps, then, the threat to the American entrepreneur is exaggerated.

A broader issue is credit availability in general. Critics of the proposed reforms charge that, in the absence of the McFadden Act, funds could be channeled too easily from one part of the country to another, leaving disadvantaged regions to atrophy while prosperous ones thrive. Testifying on Capitol Hill, the banking commissioner for the state of Texas—so far the only state to have opted out of the interstate branching provision of the Riegle-Neal Act[248]—invoked the widely feared phenomenon of "deposit siphoning":

> [U]sing our experience with interstate banking as a guide, we expect that interstate branching would accelerate the flow of credit in the economy.
>
> Nationwide networks of branches will funnel credit into growth economies and away from regions that are stagnant or experiencing a recession. This . . . activity will tend to increase the boom-bust cycle experienced in most regions and prolong economic downturns in depressed regions.[249]

Indeed, there is data to support the theory that deregulation would cause the deposits of local communities to be siphoned off to distant lending markets. In a congressional staff study of fifteen major multistate BHCs, it was determined that the BHCs' subsidiary banks—of which there were eighty-three, scattered across thirty-eight different states and the District of Columbia—frequently engaged in deposit siphoning. In particular, fully forty percent of the banks were found to have drained funds out of their host states for investment elsewhere (i.e., their in-state loan volume was significantly below average, their out-of-state loan volume significantly above average).[250] To be sure, such practices

---

[247] GAO 1994, *supra* note 58, at 16.

[248] TEX. REV. CIV. STAT. ANN. art. 489f § 2 (West 1995); *see infra* notes 371–376 and accompanying text.

[249] *Mar. 1991 Hearing*, *supra* note 147, at 487 (testimony of Kenneth Littlefield, Banking Commissioner, Texas).

[250] Analysis of Banking Industry Consolidation Issues, Staff Report to the Comm. on Banking, Finance and Urban Affairs, House of Representatives, 102d Cong., 2d Sess., Mar. 2, 1991, *in July 1993 Hearing*, *supra* note 120, at 410–11 [hereinafter 1991 Staff Report].

are economically efficient, as they allow capital to be allocated to its highest and best use.[251] But deposit siphoning runs counter to the principle that a bank should be sensitive to the needs of its community.

Community commitment, however, is not as easily abandoned as the McFadden Act's defenders suggest. Banks have two powerful incentives to plow funds back into the communities from which they were collected: it makes good business sense, and to a certain degree it is mandated by law. Quite apart from the question of civic duty, a bank has a strategic interest in investing its money at home. As one witness told the House Banking Committee: "[I]f a bank is not serving local credit needs, it is hard to see how it could remain a viable competitor on the deposit side of the balance sheet."[252]

Furthermore, in the event that the market does not provide sufficient incentives, federal law requires banks to dedicate a portion of their funds to local lending—or, more precisely, it requires that their levels of community reinvestment be taken into account as part of the regulatory examination process. Pursuant to the Community Reinvestment Act of 1977 (CRA),[253] whenever a bank applies to regulatory authorities for permission to modify or expand its business (e.g., by establishing or relocating a branch, or merging with or acquiring another institution), the authorities must consider, as a factor in their decision, the bank's "record of meeting the credit needs of its entire community, including low- and moderate-income neighborhoods."[254] Given that the regulatory system has already addressed the issue of local lending, using a statute specifically designed to police the lending markets and prevent deposit siphoning, it is unclear why further regulation is needed. It makes little sense to attack the community reinvestment problem directly, as the CRA does, and then make a second pass at it indirectly, using far-reaching prohibitions (i.e., the branching restrictions) that bear on many more questions than the narrow one at hand.[255]

This argument presupposes, of course, that the CRA is an effective regulatory tool. Although a global evaluation of the

---

[251] *See id.* at 414.

[252] *Sept. 1993 Hearing*, *supra* note 182, at 6 (testimony of E. Gerald Corrigan, chairman, Russian-American Enterprise Fund, New York).

[253] 12 U.S.C. §§ 2901–2906 (1994).

[254] 12 U.S.C. § 2903 (1994). *See* MACEY & MILLER, *supra* note 18, at 204.

[255] *See* Miller, *supra* note 32, at 1109–10.

CRA is beyond the scope of this Article, it should be noted that the statute has its detractors.[256] Describing the CRA as "ambivalent and poorly defined," one scholar criticizes it on the ground that it is only activated when a bank applies to make changes to its business beyond the ordinary course.[257] But this is precisely the appropriate moment for regulators to consider a bank's community reinvestment record, if the concern is that expansionist banks will forsake their local communities. A bank that wants to enter new markets via branching or acquisition must apply for regulatory approval, and such approval will only be granted if the institution's local lending practices are adequate.[258] In other words, it is the biggest banks—the ones with multistate ambitions—that feel the most CRA pressure.[259] The GAO cites evidence that this regulatory strategy is working: "CRA appears to have served as an incentive to spur larger banking companies that want to expand into increasing their commitment to inner cities and other underserved areas."[260] Hence the GAO's 1993 report endorses the CRA as an effective means of checking community disinvestment.[261]

More broadly, the GAO and other commentators respond to the concern over deposit siphoning by noting that the redistribution of capital, even where it involves channeling funds from one region of the country to another, is after all the point of financial intermediation. Even under the present regime, deposits may lawfully be collected in one place and loans made in another, as indeed they commonly are. The GAO remarks: "The movement of deposits from one location to make loans in another is nothing new."[262] A Treasury Department official made the same point during a congressional hearing:

> The "siphoning off" argument is no more valid for interstate branching than it is for our current system of interstate banking through holding companies. A bank that is part of an interstate holding company that does not wish to lend in an area could simply sell Federal funds upstream to a corre-

---

[256] *See, e.g.*, Wilmarth, *supra* note 9, at 1046–50.
[257] *Id.* at 1050.
[258] Miller, *supra* note 32, at 1110.
[259] *Sept. 1993 Hearing*, *supra* note 182, at 44 (testimony of Geoffrey P. Miller).
[260] GAO 1993, *supra* note 10, at 124.
[261] *Id.* at 104–08.
[262] *Id.* at 93.

> spondent bank, or could put its funds into investment securities rather than loans.[263]

Thus, it is difficult to understand why, upon repeal of the branching restrictions, banks would suddenly want to siphon funds out of their host communities when they never had before. In the GAO's view, as long as banking markets remain relatively competitive, funds will naturally flow to where they are most needed and can be of greatest use.[264] To the extent that uneconomical investment of capital is desirable—in order to provide relief to disadvantaged communities in which banks would not voluntarily choose to extend credit based on pure profit considerations—the CRA will require it.[265] The government need not impose geographic restrictions as an additional form of credit allocation.[266]

F. *Consolidation and Overconcentration*

In addition to warning against the risk of restricted access to credit, the opponents of reform argue that overconcentration would be the inevitable consequence of relaxing the geographic restrictions.[267] Indeed, in a 1993 study requested by the Senate Banking Committee, the GAO found that the consolidation trend in banking, which had its origins in the mid-1980s, would likely be accelerated by repeal of the McFadden Act and Douglas Amendment. Such acceleration was not, however, expected to be dramatic, and although consolidation would certainly produce increased concentration at both the national and regional levels, the GAO was not concerned that such concentration would be excessive.[268] With the exception of extremists like McKinsey's Lowell Bryan, who wants to see the 125 largest BHCs combined into about a dozen nationwide banks,[269] no one seriously envisions a wholesale consolidation of the banking industry.[270] In any

---

[263] *May 1991 Hearing*, *supra* note 176, at 5 (testimony of Robert R. Glauber).

[264] GAO 1993, *supra* note 10, at 108.

[265] *See* MACEY & MILLER, *supra* note 18, at 205–06.

[266] As an alternative, the geographic restrictions could be retained and the CRA discarded. Critics of the CRA prefer this approach. In any event, a single prophylactic measure—either the CRA or geographic limitations—ought to be sufficient.

[267] *See, e.g.*, Carstensen, *supra* note 90, at 1397–98.

[268] GAO 1993, *supra* note 10, at 27, 56–57.

[269] Bryan, *supra* note 172, at 374.

[270] Miller, *supra* note 32, at 1125–27 (refuting Professor Wilmarth's claim that the

event, increased concentration at the national level would not be problematic. With over 12,000 banks, the banking industry could bear a good deal of consolidation before overconcentration became an issue.[271] For that matter, the national statistics are not the important ones. Under *Philadelphia National Bank*, the relevant geographic market for purposes of antitrust analysis is the local community, not the nation as a whole. Therefore, to say that there would be a reduction in the number of banks nationwide is to say nothing.[272]

As for the effects of deregulation on local markets, fears of overconcentration appear to be unfounded. *De novo* branching would not decrease, but rather increase, the number of banks in a given market.[273] For example, a community served by eight banks under the McFadden regime might be entered by two more from out of state, for a new total of ten. If the two out-of-staters chose instead to acquire existing institutions, the total would remain eight. The market, in short, would be no more concentrated. Nor is there reason to believe that the two acquired banks would soon outperform and overwhelm the other six. In a study of 279 acquired banks during 1980-89, it was determined that if acquired banks gained any ground on their non-acquired competitors, it was only after five or more years had passed.[274] The specter of an out-of-state BHC buying up a local bank, transforming it into a powerhouse, and promptly conquering the market was judged to be unrealistic. The report concluded, "[T]his study provides little support for those who fear increasing local market dominance by interstate organizations."[275]

It does not appear that interstate acquisitions cause increased concentration at the level where it matters—the local level. A 1993 Federal Reserve study noted: "In spite of the thousands of mergers that have occurred, the average concentration in local

---

Treasury Department supported interstate banking reform in 1991 partly because it favored consolidation of the industry into a handful of megabanks).

[271] *Id.* at 1111–12.

[272] *See* Laderman & Pozdena, *supra* note 237, at 35; Savage, *supra* note 51, at 1087 ("Most bank customers are concerned with competition at the local level, rather than the national or state level.").

[273] *See* Miller, *supra* note 74, at 221 ("Establishment of a de novo bank . . . necessarily involves increased competition in local markets because the de novo bank begins with zero market share and seeks to wrest business away from existing banks in the area.").

[274] Peter S. Rose, *Interstate Banking: Performance, Market Share, and Market Concentration Issues*, 37 ANTITRUST BULL., Fall 1992, at 629.

[275] *Id.* at 630.

banking markets has remained remarkably stable over time."[276] Indeed, it is only intramarket mergers (i.e., consolidations of competitors within the same market) which reduce the number of banks in a particular community. But these mergers, of course, are not the ones implicated by interstate branching reform.

Underlying the debate over consolidation and concentration are two deeper issues: the fear that interstate branching would increase systemic risk and the fear that it would weaken competition. Some commentators have predicted that the elimination of geographic restrictions would increase the number of banks that are so large that, in the event of a crisis, the government would be forced to recognize them as being "too big to fail" (TBTF). Under the TBTF doctrine, which dates from the Continental Illinois bailout in 1984, a bank may be judged to be so central to the nation's economy that its insolvency simply cannot be tolerated, lest the banking industry and the payments system as a whole be irreparably damaged. In such cases, the FDIC satisfies the claims of not only the bank's insured depositors but also its uninsured depositors.[277] The rescue of a TBTF bank is of course extremely costly to the BIF and, ultimately, to taxpayers. As a result, Congress recently phased out the TBTF doctrine.[278]

Nevertheless, defenders of the McFadden Act warn that interstate branching might require a return to the TBTF policy. For example, the Consumer Federation of America wrote to Congress, "[B]y encouraging ever larger banks, interstate branching could well undermine the recent reforms to retire the too-big-to-fail regulatory doctrine that has ailed the federal deposit insurance system for over a decade."[279] Other commentators maintain, however, that as long as the regulatory and supervisory system continues to function correctly—particularly as fortified by the risk-based capital adequacy guidelines introduced in 1988[280] and the prompt corrective action scheme provided for in FDICIA[281]—the United States should be able to move toward a regime of

---

[276] Savage, *supra* note 51, at 1089.

[277] *See* GAO 1993, *supra* note 10, at 85–86; Wilmarth, *supra* note 9, at 994–1003.

[278] Section 141 of FDICIA prohibits TBTF bailouts after December 31, 1994, except in extraordinary circumstances. 12 U.S.C. § 1823(c)(4)(E)(i) (1994).

[279] *July 1993 Hearing*, *supra* note 120, at 98 (statement of the Consumer Federation of America).

[280] 12 C.F.R. § 3 (1995).

[281] Section 131 of FDICIA. 12 U.S.C. § 1831o (1994).

interstate branching without having to re-adopt the TBTF policy.[282]

Even more importantly, critics of interstate branching allege that it would undermine competition and, thus, raise the price and decrease the quality of banking services available to the public.[283] Fears of diminished competition do not, however, appear to be warranted. Concentration is only dangerous to the extent that it leads to oligopolistic behavior; it poses no threat so long as the industry remains competitive, as it appears the banking industry would. For example, the GAO concluded that, far from reducing competition, deregulation would actually enhance it: new players would enter once-protected markets, and established banks would tend to lower prices and improve services in order to prevent their customers from defecting to out-of-town institutions.[284] In fact, the mere potential for entry—even if out-of-state banks were not to exercise their option to expand immediately—would likely increase the competitive pressure on local banks.[285]

These theoretical assertions appear to be borne out by reality. A study by the Federal Reserve Bank of San Francisco determined that relaxation of geographic restrictions "tends significantly to enhance potential and/or actual competition in state banking markets."[286] On the same note, the GAO's 1993 report indicated that there was no evidence that banks belonging to multistate BHC families had engaged in uncompetitive behavior, such as predatory pricing.[287] In sum, it appears unlikely that deregulation would reduce competition, as its opponents maintain.

In addition to the market forces keeping the banking industry competitive and unconcentrated, there are also federal and state statutory schemes designed to guard against monopolistic and oligopolistic behavior.[288] *Philadelphia National Bank* established that the banking industry is subject to the antitrust laws just like any other industry.[289] In addition, banks must seek pre-approval

---

[282] *See, e.g.*, Miller, *supra* note 32, at 1106–08.
[283] *See, e.g.*, Carstensen, *supra* note 90, at 1397–98.
[284] GAO 1993, *supra* note 10, at 118.
[285] Carstensen, *supra* note 90, at 1419.
[286] Laderman & Pozdena, *supra* note 237, at 43.
[287] GAO 1993, *supra* note 10, at 117.
[288] *See* MACEY & MILLER, *supra* note 18, at 442; GAO 1993, *supra* note 10, at 126–48.
[289] 374 U.S. 321 (1963).

from federal authorities in order to carry out a merger.[290] A bank's merger application will be denied if the proposed transaction would result in a monopoly or conspiracy to monopolize (pursuant to the Sherman Act),[291] or if it would have the effect of substantially lessening competition (pursuant to the Clayton Act).[292] In addition, banks are subject to special, industry-specific merger statutes, most notably the Bank Merger Act,[293] which covers mergers between federally insured commercial banks, and the BHCA,[294] which covers BHC consolidations. Given that the various states also have antitrust statutes, it is clear that there is no dearth of statutes aimed at preventing predatory pricing, excessive market power, and other anticompetitive behavior. Some commentators, of course, have less faith than others in the ability of the antitrust laws to police the market.[295] But it hardly seems necessary to supplement the antitrust laws, which are specifically targeted at the problem of uncompetitiveness, with general restrictions on bank expansion, which sweep much more broadly—and probably overbroadly.[296] To the extent that regulation is needed to protect competition, existing antitrust laws—provided they are vigorously enforced—should be able to do the job.[297]

One source of competition for large banking organizations is the community banks, which have succeeded in carving out a niche for themselves even in markets where they are no longer protected by branching restrictions.[298] The evidence belies the argument, advanced by opponents of deregulation, that interstate branching would result in the disappearance of community banks.[299] On the contrary, they are likely to continue to prosper. Although large banks have certain advantages—they can invest

---

[290] Bank Merger Act, 12 U.S.C. § 1828(c) (1994).

[291] 15 U.S.C. § 2 (1994).

[292] 15 U.S.C. § 18 (1994).

[293] 12 U.S.C. § 1828(c) (1994).

[294] 12 U.S.C. § 1842(c) (1994).

[295] Professors Carstensen and Wilmarth argue that the current antitrust regime for banks is inadequate. *See* Carstensen, *supra* note 90, at 1436; Wilmarth, *supra* note 9, at 1024–27.

[296] *See* Miller, *supra* note 32, at 1117. Professor Miller makes the same argument concerning the regulation of community reinvestment. *See supra* notes 265–266 and accompanying text.

[297] *See* Cohen 1985, *supra* note 38, at 977; GAO 1993, *supra* note 10, at 147–48; Miller, *supra* note 32, at 1111.

[298] *See* Miller, *supra* note 32, at 1114.

[299] *See, e.g.*, Wilmarth, *supra* note 9, at 1040, 1042–43.

heavily in marketing and advertising,[300] display a familiar (and perhaps trusted) name,[301] and offer an exceptionally wide range of services[302]—small banks also have unique qualities that make them attractive to a substantial segment of the market. Small banks customarily know their communities better and are more responsive to local needs, or at least they portray themselves in this light in order to win consumer loyalty.[303] The president of a small Long Island bank, testifying before Congress, described the independent bankers' philosophy and explained why he felt their position was secure:

> As a community banker, it is our goal not to be all things to all people, but rather to know our customers and their businesses so that we can give them personalized services and specialized products . . . . Due to the unique role of a community bank, I have not felt the negative impact of consolidation, and do not believe that further consolidations, which would increase the presence of larger banks in our area, will negatively affect our financed [sic] growth and success.[304]

Empirical evidence demonstrates that this view is sound. The GAO's 1993 report on interstate banking, among other recent studies,[305] documents the success of the independent banks and predicts that, even under a regime of interstate branching, thousands of small institutions would survive and their market shares would be undiminished.[306] Even extremists like Bryan, who wants the 125 largest BHCs to be collapsed into ten to fifteen nationwide megabanks, see a place for community banks in the new regime: "This consolidation would not significantly affect the

---

[300] In a prepared statement submitted to the House Banking Committee, the Consumer Federation of America lamented the big banks' marketing power: "One would think that if smaller banks offered better services at a more competitive price, consumers would be flocking to the smaller institutions. Unfortunately, this is not the case . . . . The powerful marketing forces of the larger institutions overwhelm consumers. Consumers are often unaware of alternative and more competitive market products." *July 1993 Hearing*, *supra* note 120, at 112 (statement of the Consumer Federation of America).

[301] *See* Freeman, *supra* note 96, at 1149.

[302] *See* GAO 1993, *supra* note 10, at 103; Bill Streeter & Steve Cocheo, *Merge or Die?*, ABA BANKING J., Aug. 1 1994, at 103 ("Size means more than efficiency. It also means market presence, customer awareness, and marketing clout.").

[303] Community banks have been compared to micro-breweries and mini-mills. *See* Marilyn R. Seyman & Steven P. Williams, *To Stay Independent, Check Cost Structures*, AM. BANKER, Jan. 6, 1995, at 9, *available in* LEXIS, Bankng Library, Ambank File.

[304] *July 1993 Hearing*, *supra* note 120, at 12–13 (testimony of Paul C. Settelmeyer, president, Bank of Great Neck).

[305] *See, e.g.*, Rose, *supra* note 274, at 630; Savage, *supra* note 51, at 1081–82.

[306] GAO 1993, *supra* note 10, at 66.

thousands of small, profitable local banks with deep roots (and thus real competitive advantage) in their communities."[307]

G. *Historical Experience with Intrastate Branching and Interstate Banking*

Some large states have abandoned branching restrictions in favor of statewide branching. These states may be viewed as laboratories in which deregulation experiments have already been conducted. Their experiences suggest that small, independent banks can and would survive under an interstate branching regime. New York, for example, authorized statewide branching in 1960. Since that time, its banking markets have not been dominated, as many feared they would be,[308] by money-center banks from Manhattan. On the contrary, in the decades that followed deregulation, the total market share of the three largest banks remained essentially constant.[309] The money-center banks that branched upstate were unable to commandeer the markets there. As the New York Superintendent of Banks told Congress:

> The fear that expanding branching authority would pose a threat to small independent banks is understandable, although in New York we have found that fear to be largely without rational foundation . . . . Indeed, our experience in New York with both statewide intrastate branching as well as interstate bank holding company expansions has proved to us that small local banks can continue to thrive under interstate branching.[310]

The case of California is somewhat less clear. California was among the first states to permit statewide branching; its statute dates from 1909. The California banking industry is highly concentrated and, by some accounts, unresponsive to consumers' needs. Three institutions—Bank of America, Wells Fargo, and First Interstate—hold more than half of the state's deposits, and banking services are higher-priced than the national average. California banks charge more for loans and pay less for deposits than banks in other states.[311] Although out-of-state entry has

---

[307] Bryan, *supra* note 172, at 74.

[308] *See Banking Industry Made More Progress in Risk Management, McDonough Says*, 64 BNA'S BANKING REP. 274, 274 (1995).

[309] Cohen 1985, *supra* note 38, at 979–80.

[310] *July 1993 Hearing*, *supra* note 120, at 9 (testimony of Derrick D. Cephas, Superintendent of Banks, New York).

[311] *See id.* at 111 (statement of the Consumer Federation of America); Wilmarth, *supra* note 9, at 1022.

been authorized since 1987, out-of-state banks have been unable to compete in California markets and today hold less than one percent of the state's deposits.[312] Given the size and scope of California's economy, it is probably the most useful laboratory in the country for predicting the economic effects of a nationwide branching regime. Hence these results are alarming.

There is no consensus, however, about how to interpret the California experience. A Treasury Department official urged Congress to repeal the McFadden Act on account of what he perceived to be the California success story:

> Take a State like California, which, by all measures, I would say is like a country in terms of its size. We have what amounts to nationwide branching within California. And what has been the effect? Has it driven all the independent bankers out of business in California as Bank of America has opened up a storefront branch across the street? The answer is no. We have an incredibly rich, diverse, competitive banking environment in California with both branches of large banks and very, very healthy independent banks.[313]

This optimistic view, expressed in 1991, is substantiated by recent data from the GAO, which published a California case study in 1994. Based on data for the period 1984 to 1993, the GAO found that despite the ubiquitousness of the three megabanks—with their statewide branching networks—small and medium-sized banks had fared extremely well and were "a profitable and viable part of California's banking industry."[314] In the GAO's view, it is fair to generalize from the California experience that independent banks can coexist alongside large institutions spread out over vast areas.[315]

This generalization is also borne out by data from Arizona and Washington, two additional states considered in the GAO's 1994 report. Once Arizona and Washington enacted interstate banking statutes,[316] effective in 1986 and 1987 respectively, out-of-state institutions (particularly from California) entered their markets and established strong positions. The independent banks in Washington, however, did not see their market share diminished, and

---

[312] GAO 1994, *supra* note 58, at 6, 32–33.

[313] *May 1991 Hearing*, *supra* note 176, at 10 (testimony of Robert R. Glauber, Under Secretary (Finance), U.S. Dep't of the Treasury).

[314] GAO 1994, *supra* note 58, at 41, 44.

[315] *Id.* at 12.

[316] ARIZ. REV. STAT. ANN. § 6-321 (1989); WASH. REV. CODE ANN. § 30.04.232 (West 1986 & Supp. 1995).

although Arizona's small banks suffered during the recession of the late 1980s, they regained strength after 1990 and have reestablished themselves as a viable force in the industry.[317] In light of this evidence, the GAO concluded:

> Although states' interstate banking and in-state branching laws provided large banks with the opportunity to expand, the experiences of California, Washington, and Arizona indicated that such geographic deregulation did not necessarily result in a more concentrated industry . . . . [S]maller banks . . . continued to play an important role. They frequently were among the most profitable banks as measured by return on assets and, despite geographic deregulation, either gained additional market share or regained previously lost market share.[318]

In short, if the experience of the states is any indication, there is little reason to expect that liberalization of the branching laws would leave banking markets uncompetitive or vulnerable to abuse on the part of large institutions.

H. *Foreign Banking Systems*

It was often observed during the deregulation debate that countries do not cabin their banks within narrow boundaries the way the United States does.[319] The Canadian banking industry, for example, is dominated by a few large institutions that operate nationwide branching networks.[320] A Quebecer can walk into a Royal Bank of Canada branch office in Vancouver and transact business as if he had never left Quebec.[321] The notion of a bank's commanding one market and one market only—as, for example, First Chicago does in the greater Chicago area—is unknown in Canada. The Bank of Montreal serves the entire country despite its name. The same principles hold true for Japan and Western Europe. In Japan, a handful of large banks called "city banks" maintain branches throughout the country and hold about half the nation's deposits.[322] In Europe, banks are free to branch

---

[317] GAO 1994, *supra* note 58, at 6, 56, 58, 80.
[318] *Id.* at 4–5.
[319] GAO 1993, *supra* note *ff*, at 23.
[320] MACEY & MILLER, *supra* note 18, at 15, 390.
[321] Sept. 1993 Hearing, *supra* note 182, at 43 (testimony of Geoffrey P. Miller).
[322] MACEY & MILLER, *supra* note 18, at 390; Bill Shaw & John R. Rowlett, *Reforming the U.S. Banking System: Lessons from Abroad*, 19 N.C. J. INT'L L. & COM. REG. 91, 105 (1993).

nationwide and even across national borders. Pursuant to the Second Banking Directive of 1989, effective January 1, 1993, a bank chartered in any European Community (EC) member state may establish a branch in any other EC country.[323] On account of this freedom to expand, BHCs are unneeded in Canada, Japan, Great Britain, Germany, and France.[324] Moreover, compared to other industrialized nations, the United States has an extraordinarily diffuse banking system.[325] Whereas there are over 12,000 commercial banks in the United States, there are 65 in Canada, 150 in Japan, 550 in Britain, 900 in Germany, and 400 in France.[326] Only two American banks, Citibank and the combined Chemical Bank–Chase Manhattan are among the top fifty banks worldwide.[327] Clearly, then, the U.S. banking system is an aberration.

From this fact can be drawn either of two conflicting conclusions. The prevailing view is that it is harmful to America's competitiveness to be the only major nation in the world that corrals its banks in this manner. Forced to comply with rules that limit their profitability and efficiency, American banks cannot perform at the level of their foreign counterparts.[328] Senator Riegle argued to his colleagues:

> The U.S. is the only industrial country that restricts bank branching. The globalization of the banking industry means that U.S. banks cannot afford to continue to base their success on a limited geographic area. They cannot match their competitors while burdened with costly subsidiary structures and cannot be strong global competitors without larger deposit bases in this country.[329]

Moreover, quite apart from the question of global competitiveness, it is worth asking what justification the United States has for imposing geographic restrictions when the rest of the world apparently sees no need for them. The defenders of the old order have the burden of proving that every other industrialized democracy in the world has it wrong. Some evidence suggests that

---

[323] Second Banking Directive, 89/646/EEC (1989). *See infra* notes 414–421 and accompanying text.

[324] Melanie L. Fein, *Interstate Branching: A Formula for Refashioning Supervision and Regulation*, BANKING POL'Y REP., Aug. 15, 1994, at 15.

[325] GAO 1993, *supra* note 10, at 29.

[326] Miller, *supra* note 32, at 1122.

[327] *Id.* at 1122; *Two New Mergers Create Global, Regional Banks*, S.F. CHRON., Aug. 29, 1995, at B8.

[328] *See, e.g.*, MODERNIZING THE FINANCIAL SYSTEM, *supra* note 71 at x, 7–8, 55–57; GAO 1993, *supra* note 10, at 23–24.

[329] 140 CONG. REC. S4765 (daily ed. Apr. 25, 1994) (statement of Senator Riegle).

other banking systems, because they are so highly concentrated, may not be as competitive and may not provide the same level of service as the American system.[330] In that case, the United States may be better for its strangeness. But the opposing view—that the United States should get in step with the rest of the world—appears to have carried the day, and probably rightly so.

I. *Tax Revenues*

Finally, opponents of interstate branching have expressed fears that it may cause the states to lose tax revenue. In its narrowest expression, the argument is that if, in a post-McFadden world, a bank headquartered in State *A* established a branch in State *B*, State *B* would not have the power under 31 U.S.C. § 3124 to tax the branch's income from federal obligations because State *B* would be neither the branch's state of domicile nor the grantor of the privilege to branch interstate.[331]

Not all commentators, however, have limited themselves to such carefully circumscribed—or highly technical—arguments. Rather, a much broader assertion has often been made: if a bank subsidiary of an out-of-state BHC is converted into a branch, the host state will lose its taxing jurisdiction altogether. The Independent Bankers Association of America (IBAA), which represents over 5600 small banks, has urged its members to protest interstate branching on this ground:

> Today, virtually every state taxes banks chartered by or headquartered in the state based on the fact that the bank has a corporate entity in the state. If a bank becomes merely a branch of an out-of-state bank, there will no longer be a corporate entity in your state, and the tax laws will not apply.[332]

---

[330] For example, Professor Wilmarth argues that the British and Canadian banking markets are "characterized by oligopolistic behavior." He points to the "demonstrably inferior services these banks provide to consumers and small businesses when compared to U.S. banks." Wilmarth, *supra* note *ff*, at 1052–53.

[331] SANDRA B. MCCRAY, THE EFFECT OF THE REPEAL OF THE MCFADDEN ACT ON STATE TAX REVENUE (1991), *in Impact of Bank Reform Proposals on Consumers: Hearing before the Subcomm. on Consumer Affairs and Coinage of the House Comm. on Banking, Finance and Urban Affairs*, 102d Cong., 1st Sess. 503–18 (1991).

[332] Independent Bankers Association of America, Interstate Banking (unpublished and undated manuscript, on file with author).

The solution is simply for the states to modify their tax codes so that their taxing jurisdiction is triggered by something less than the presence of a full-fledged "corporate entity." A more sensitive trigger is entirely permissible; it is well established that a state may tax a branch of an out-of-state bank.[333] A Treasury Department official reassured Congress, "States will have the same authority to raise revenue from bank branches as they do from banks, businesses, or local franchises."[334] Consequently, concerns for the tax collector are misplaced.

J. *Summary*

These were the points and counterpoints raised during the 1993–94 debate over interstate banking and branching. Many of the arguments had been in the air since the early- to mid-1980s. For that matter, many had been raised as early as the 1920s, when the McFadden Act was originally under consideration.[335] In this latest round of discussions, however, the free market arguments had more force than ever before. Those calling for deregulation were now in the majority and, as this Part has shown, their views made sense for American banks on the eve of the twenty-first century. Although the defenders of the old order had tradition on their side, the reformers generally had the more compelling data on theirs. As John LaWare, Federal Reserve governor, observed: "We believe that no good public policy purpose is served by restraining the freedom of choice of individual banking organizations whose managers know best what is the least-cost operating structure for them."[336] In the interest of efficiency, convenience, competition, and safety and soundness, it was time to demolish the walls separating the different regions of the country and allow banks to determine their own fields of operation in accordance with market principles.

---

[333] *See* Sandra B. McCray, *State Taxation of Interstate Banking*, 21 GA. L. REV. 283, 297 (1986); Northwestern States Portland Cement Co. v. Minnesota, 358 U.S. 450 (1959) (holding that a state has jurisdiction to tax a nonresident entity that owns or leases an in-state office). For an excellent survey of the tax issues, see Henry Ruempler, *Will State Taxation Tail Ultimately Wag Interstate Banking Dog?*, BANKING POL'Y REP., Jan. 16, 1995, at 1.

[334] *May 1991 Hearing*, *supra* note 176, at 5 (testimony of Robert R. Glauber).

[335] *See* Nestor, *supra* note 42, at 612–13.

[336] *June 1993 Hearing*, *supra* note 64, at 5 (testimony of John P. LaWare, governor, Board of Governors of the Federal Reserve System).

IV. The New Political Compromise

A. *The Triumph of the Free Market*

The legislation that emerged from this debate, the Riegle-Neal Act of 1994, is a triumph for proponents of the free market. The new law repeals the two principal statutes blocking geographic expansion by banks, the Douglas Amendment and the McFadden Act, effective mid-1995 and mid-1997, respectively. Thus, it is no longer possible for a state to prohibit acquisition of in-state banks by out-of-state BHCs, or to permit only BHCs from certain neighboring states to enter, while those from distant states or from the so-called money centers are excluded. In an even more dramatic step, the Riegle-Neal Act authorizes BHCs to combine their subsidiary banks in different states into a single multistate bank. The implications of this measure are startling. A single bank—having only one charter, one board of directors, and one management—will be able to establish branches nationwide, offering customers from coast to coast exactly the same products and services, packaged and delivered in exactly the same way. Thus, the Riegle-Neal Act stands fundamentally for the proposition that banks should be able to take their business where they like. Rather than having to respect artificial political boundaries, they will henceforth be encouraged to obey only their own instincts.

B. *The Preservation of States' Rights*

In addition to this free market notion, there is a second major theme underlying the Riegle-Neal Act. By deferring to the states on a wide variety of issues, ranging from concentration limits to consumer protection, Congress repeated essentially what President Jackson had stated in vetoing the recharter of the Second Bank of the United States in 1832[337]—the federal government must not trespass upon areas reserved for regulation by the states. In the Joint Explanatory Statement accompanying the new legislation, the conference committee made precisely this point:

> States have a strong interest in the activities and operations of depository institutions doing business within their juris-

---

[337] *See supra* notes 21–22 and accompanying text.

> dictions, regardless of the type of charter an institution holds. In particular, States have a legitimate interest in protecting the rights of their consumers, businesses, and communities. Federal banking agencies, through their opinion letters and interpretive rules on preemption issues, play an important role in maintaining the balance of Federal and State law under the dual banking system. Congress does not intend that the Interstate Banking and Branching Efficiency Act of 1994 alter this balance and thereby weaken States' authority to protect the interests of their consumers, businesses, or communities.[338]

The Riegle-Neal Act, then, is as much a states' rights statute as it is a free market statute. As will be discussed, at its most extreme the new legislation authorizes states to remove themselves from the statutory framework altogether, at least with respect to interstate branching. But on a more modest scale, the Riegle-Neal Act also carves out a number of zones where the states retain exclusive regulatory authority, or where federal law only applies in the absence of state law on the same subject. In short, the statute is pervaded with a sense that states' rights must be respected.

In the section that repeals the Douglas Amendment, reference is repeatedly made to the laws of the relevant home state (i.e., the state in which the national bank has its main office).[339] Although states cannot opt out of the interstate banking provision, they are nevertheless allowed to dictate many of the terms under which out-of-state BHCs may acquire banks within their boundaries. State age laws, for example, are not preempted by the federal statute: a state may limit the pool of potential targets to those banks that have been in existence for a certain period, up to five years.[340] It may require the resulting bank to set aside part of its assets for public housing.[341] It may also vary the concentration limits fixed by the Riegle-Neal Act, which provides that an acquisition may not be approved if the resulting bank would

---

[338] 140 CONG. REC. H6638 (Aug. 2, 1994) (Joint Exmplanatory Statement of the Comm. of Conference) [hereinafter Joint Explanatory Statement].

[339] "The term 'home State' means . . . with respect to a national bank, the State in which the main office of the bank is located." Riegle-Neal Act, § 101(c) (amending Bank Holding Company Act § 2(o)(4)).

[340] *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(1)(B)).

[341] "No provision of this subsection shall be construed as affecting the applicability of a State law that makes an acquisition of a bank contingent upon a requirement to hold a portion of such bank's assets available for call by a State-sponsored housing entity established pursuant to State law." *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(1)(D)).

control more than ten percent of insured deposits nationwide[342] or more than thirty percent of insured deposits in-state.[343] States are free to raise or lower this thirty percent cap.[344] In addition, the Riegle-Neal Act is careful not to interfere with other forms of state regulation. The statute explicitly provides that state community reinvestment and antitrust laws, as well as their federal counterparts, are to remain undisturbed,[345] and that state taxation authority is undiminished.[346]

Similarly, the branching section of the Riegle-Neal Act grants broad discretion to the states. The statute repeats the earlier provisions concerning concentration limits, community reinvestment, antitrust, tax, and age laws,[347] and further provides that out-of-state branches may be regulated by the host state. Thus, when a bank headquartered in one state branches into another, it subjects its branches to that state's regulatory regime: "The laws of the host State regarding community reinvestment, consumer protection, fair lending, and establishment of intrastate branches shall apply to any branch in the host State of an out-of-state national bank to the same extent as such State laws apply to a branch of a bank chartered by that State."[348] The host state is also entitled to decide whether out-of-state banks may enter via *de novo* branching,[349] or whether they must first acquire

---

[342] *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(2)(A)).

[343] *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(2)(B)).

[344] "No provision of this subsection shall be construed as affecting the authority of any State to limit, by statute, regulation, or order, the percentage of the total amount of deposits of insured depository institutions in the State which may be held or controlled by any bank or bank holding company (including all insured depository institutions which are affiliates of the bank or bank holding company)." *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(2)(C)).

[345] *Id.* § 101(a) (amending Bank Holding Company Act § 3(d)(3)–(4)).

[346]

> No provision of this Act shall be construed as affecting the authority of any State or political subdivision of any State to adopt, apply, or administer any tax or method of taxation to any bank, bank holding company, or foreign bank, or any affiliate of any bank, bank holding company, or foreign bank, to the extent that such tax or tax method is otherwise permissible by or under the Constitution of the United States or other Federal Law.

Id. § 101(b) (amending Bank Holding Company Act § 7(b)).

[347] *Id.* § 102(a) (amending Federal Deposit Insurance Act §§ 44(a)(5), 44(b)–(c)).

[348] *Id.* § 102(b) (amending 12 U.S.C. § 36(f)(1)(A)).

[349]

> The term "de novo branch" means a branch of a national bank which (i) is originally established by the national bank as a branch; and (ii) does not become a branch of such bank as a result of (I) the acquisition by the bank of an insured depository institution or a branch of an insured depository institution; or (II) the conversion, merger, or consolidation of any such institution or branch.

Id. § 103(a) (amending 12 U.S.C. § 36(g)(3)(A)).

existing banks and then convert them into branches.[350] The distinction is important because a bank's stock usually rises when it becomes a potential acquisition target.[351] Consequently, small banks may be inclined to lobby for an acquisition-only rule in order to enhance the franchise value of their charters.[352] In the event of acquisition, the host state may specify whether the out-of-state institution must acquire an entire bank, or may instead simply acquire a branch and disregard the rest.[353] The host state may also impose reasonable notification and reporting requirements on branches of out-of-state banks.[354] Finally, states retain full authority to grant or deny interstate branching privileges to their own state-chartered institutions.[355] In these ways, the Riegle-Neal Act demonstrates that Congress remains faithful to the principles of Jacksonian federalism in the bank regulatory field.

---

[350] *Id.* § 103(a) (amending 12 U.S.C. § 36(g)(1)(A)). In the language of the statute, "opt-in" denotes a state's election to allow *de novo* branching. In popular parlance, however, commentators speaking of the "opt-in" provision of the Riegle-Neal Act mean the provision that allows states to accelerate the effective date of the interstate branching rule (June 1, 1997), thereby permitting national banks to get an early start on branching across state lines. *See id.* § 102(a) (amending Federal Deposit Insurance Act § 44(a)(3)). The dual use of the term "opt-in"—one technically correct, the other commonly accepted—may cause confusion. This Article uses the term "early-in," rather than "opt-in," to refer to a state's election to permit interstate branching prior to the effective date.

[351] *See* MACEY & MILLER, *supra* note 18, at 430.

[352] *See* Savage, *supra* note 51, at 31. Small banks face a difficult dilemma, however, because a *de novo* branching rule could help them as well. Opening a new branch from scratch is less expensive than acquiring an existing franchise and converting it into a branch. Under an acquisition-only rule, only big banks would be able to afford to expand across state lines. By contrast, if *de novo* branching were permitted, small banks would also be able to participate in interstate branching. For a discussion of the *de novo* problem, see Bryan Walpert, *Erasing Invisible Lines That Hem Banks Could Put Md. Assembly on the Spot*, WARFIELD'S BUS. REC., Nov. 18, 1994, at 5 (interview with John B. Bowers, Jr., executive vice president, Maryland Bankers Association) [hereinafter Bowers Interview].

[353]

> An interstate merger transaction may involve the acquisition of a branch of an insured bank without the acquisition of the bank only if the law of the State in which the branch is located permits out-of-State banks to acquire a branch of a bank in such State without acquiring the bank.

Riegle-Neal Act § 102(a) (amending Federal Deposit Insurance Act § 44(a)(4)).

[354] *Id.* § 102(a) (amending Federal Deposit Insurance Act § 44(c)(4)).

[355] "No provision of this section shall be construed as limiting in any way the right of a State to (A) determine the authority of State banks chartered by that State to establish and maintain branches." *Id.* § 102(a) (amending Federal Deposit Insurance Act § 44(c)(3)). As a practical matter, most states will probably grant interstate branching privileges to their state-chartered banks so as not to put them at a competitive disadvantage. Otherwise banking organizations would have an incentive to convert to national charters, causing revenue loss to the states. *See State Action Necessary to Let State Banks Branch Interstate, CSBS Advises*, BANKING POL'Y REP., Nov. 7, 1994, at 2.

The enduring strength of the states' rights doctrine is particularly apparent in the provisions allowing states to escape from the interstate branching scheme altogether, or alternatively to "early-in." The Riegle-Neal Act only provides for interstate branching as the background rule; the states may opt out.[356] In other words, a state may elect to keep the McFadden Act alive by requiring that any out-of-state banking organization seeking entry must utilize the BHC format, with separately chartered banks, rather than the multistate branch network structure. Hence, the Riegle-Neal Act comes close to "Douglasizing" the McFadden Act; it lodges in the states all final authority to determine the extent of branching, on the part of both state and national banks.[357] The only difference is in the default rule. Whereas the Douglas Amendment permitted states to opt in to interstate banking, the Riegle-Neal Act makes interstate branching the default regime and allows states to opt out.

The commentators are split over whether "Douglasizing" (or, more accurately, semi-"Douglasizing") the McFadden Act is a wise idea. Professor Arthur Wilmarth, one of the most outspoken critics of interstate branching proposals, maintains that the background rule should be a ban on interstate branching, and that if any reform is to be undertaken, it should consist at most of limited, state-by-state experimentation with lifting this ban.[358] Testifying before a House subcommittee, Professor Wilmarth opined that Congress "should allow the States to retain a sig-

---

[356] "[A] merger transaction may not be approved . . . if the transaction involves a bank the home State of which has enacted a law after the date of enactment of the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 and before June 1, 1997, that . . .expressly prohibits merger transactions involving out-of-State banks." Riegle-Neal Act § 102 (amending Federal Deposit Insurance Act § 44(a)(2)(A)). It is not uncommon for Congress to include an opt-out clause in a statute in order to ensure that state law will not be preempted against a state's will. One precedent for use of the opt-out device in the area of banking law is Section 501 of the Depository Institution Deregulation and Monetary Control Act of 1980 (DIDMCA), 12 U.S.C. § 1735f-7 (1994), which preempted state usury limits on residential mortgages but allowed states to override federal preemption. *See* MACEY & MILLER, *supra* note 18, at 125. Curiously, the opt-out mechanism was also used in the 1980s at the state level to allow *banks* to remove themselves from the reach of state interstate banking statutes. Indiana and Ohio, for example, permitted banks to opt out of their interstate banking schemes. A bank that opted out could be neither a target nor an acquirer in any interstate acquisition. *See* MCCOY ET AL., *supra* note 51, at 113.

[357] The states were always free to permit interstate branching by their own state-chartered institutions. Indeed, before passage of the Riegle-Neal Act, eight states had enacted interstate branching statutes. GAO 1994, *supra* note 58, at 3. The Riegle-Neal Act gives national banks the same interstate branching privileges, but allows states to suspend those privileges for national banks located within their boundaries.

[358] Wilmarth, *supra* note 9, at 1074.

nificant role in regulating entry by big banks . . . . [A] State legislature is in a better position than Congress to decide what type of banking structure is best for its citizens and businesses."[359] On the other hand, some feel that merely "Douglasizing" the McFadden Act is too meager a step in the direction of full nationwide banking. Fervent reformers like Professor Geoffrey Miller view this simply as a "halfway measure," and argue that the states should not be allowed to veto selectively the congressional initiative.[360]

C. *The Choice to Opt Out*

The decision to forward the branching question to the states means that the Riegle-Neal Act is really only a political compromise. Congress has not spoken definitively or authoritatively on the issue of interstate branching. Rather, it has shifted the deregulation debate from Washington, D.C., to state capitals across the country. In any given state, the shape of the debate depends upon that state's present banking climate.

In Colorado, for example, tremendous resources have been expended on both sides of the opt-out debate. In February 1995, the Colorado legislature became the first in the country to pass an opt-out bill, only to see it vetoed by Governor Roy Romer.[361] Colorado was a likely opt-out candidate at the time because its five largest banks were owned by out-of-state BHCs,[362] and it had no megabanks or superregionals of its own.[363] Consequently, only banks owned by out-of-state organizations, such as Norwest and First Bank Systems of Minnesota, lobbied for interstate branching. The legislature chose instead to listen to the independent banks—all Colorado-owned—that demanded that their markets be protected from out-of-state entry. The community

---

[359] *See Sept. 1993 Hearing*, *supra* note 182, at 55 (testimony of Arthur E. Wilmarth).

[360] Miller, *supra* note 32, at 1124–25.

[361] *See Colorado Legislature Votes to Opt Out of Riegle-Neal Scheme*, BNA's BANKING REP., Feb. 27, 1995, at 424; Terrence O'Hara, *Governor Slaps a Veto on Colorado Opt-Out Bill*, AM. BANKER, Mar. 14, 1995, at 1 [hereinafter O'Hara, *Governor*], *available in* LEXIS, Bankng Library, Ambank File; *Colorado Governor Vetoes Measure to "Opt Out" of Branch Banking*, DAILY REP. FOR EXECUTIVES, Mar. 13, 1995, at 48, *available in* LEXIS, Legis Library, Drexec File.

[362] Norwest, First Bank Systems, First Interstate, BancOne, and Key Banking Corp together hold 68% of the state's deposits. *See* O'Hara, *Governor*, *supra* note 361, at 1.

[363] The term "superregional" refers to a large multistate bank not headquartered in a money center. Among the most successful superregionals today are NationsBank, of North Carolina, and BancOne, of Ohio.

banks have traditionally been a powerful force in Colorado. Largely on account of their lobbying efforts, Colorado remained a unit-banking state longer than any other state in the nation, only adopting in-state branching in 1991.[364] Consumers also supported the opt-out measure, as noted by a representative of the Independent Bankers of Colorado:

> [T]he reason this bill was passed so overwhelmingly was because of the way those [out-of-state] bank holding companies have treated Colorado bank customers . . . . When they passed interstate banking years ago, all the big banks promised that Denver would be the biggest financial center between St. Louis and San Francisco. Well, they've listened to their constituents complain about called loans, loans not being renewed, and impersonal service from someone in Minneapolis and, well, it added up.[365]

Thus, the Colorado legislature decided that the appropriate course for Colorado was to opt out of the interstate branching component of the Riegle-Neal Act.

Governor Romer, however, recognized the danger of standing aside while the rest of the nation went forward with interstate branching. A state that opts out could find itself at a competitive disadvantage because, as a concomitant to its prohibiting entry by out-of-state banks, its own state- and federally chartered banks are forbidden to branch across state lines.[366] States have an incentive to keep pace with one another, in order to attract capital and employers.[367] By preserving archaic geographic restrictions, a state is likely to drive away banking business. In addition, there is some risk that if a state opts out, it might be barred from ever opting back in. A number of law firms have given opinions to this effect.[368] Although a state's ability to opt back in will

---

[364] Ellen Miller & Henry Dubroff, *War Pits Big Banks vs. Small*, DENVER POST, Dec. 24, 1994, at A1; Barbara A. Rehm, *Colorado Ready to Finally Allow Branch Banking*, AM. BANKER, May 10, 1991, at 2, *available in* LEXIS, Bankng Library, Ambank File.

[365] James Thomas, executive manager, Independent Bankers of Colorado, *quoted in* Terrence O'Hara, *A Colorado Thumbs Down for Branching*, AM. BANKER, Feb. 24, 1995, at 1, *available in* LEXIS, Bankng Library, Ambank File.

[366] "A bank whose home State opts out of interstate branching may not participate in any interstate merger transaction." Joint Explanatory Statement, *supra* note 338, at H6639.

[367] *See* Savage, *supra* note 51, at 27–29.

[368] The firms are Day, Edwards, Federman, Propeater & Christiansen, of Oklahoma City, Okla.; Foss & Moore, of Bismarck, N.D.; and Barnett & Sivon, of Washington, D.C. A contrary opinion has been provided to the Conference of State Bank Supervisors by Arthur Wilmarth, in his capacity as partner of Barley, Snyder, Senft & Cohen, of Lancaster, Pennsylvania. *See* Terrence O'Hara, *Pro-Interstate Okla. Banks Say Holdout States Might Be Stuck with the Decision*, AM. BANKER, Mar. 1, 1995, at 8, *available*

ultimately have to be decided by a federal court—which is unlikely to support a bar against opting back in, in light of the Act's legislative history and intent—the three legal opinions in question have justifiably created concern. Thus, although the IBAA and other supporters of Colorado's opt-out bill accused Governor Romer of putting national and out-of-state interests ahead of Colorado's own,[369] he acted reasonably to ensure that his state would be included in the new nationwide banking and branching system. The legislature subsequently enacted, and Governor Romer signed, compromise legislation that, while it does not expressly authorize interstate branching, acquiesces to it as of the federal trigger date (June 1, 1997).[370]

Although the opt-out movement lost in Colorado, it carried the day in Texas. The nation's first (and to date its only) opt-out statute was signed into law by Texas Governor George Bush on May 10, 1995.[371] Texas's resistance to interstate branching stems from the fact that it is heavily populated by independent banks and lacks major money-center or superregional institutions.[372] There is no Texas equivalent of California's BankAmerica or North Carolina's NationsBank. Rather, a handful of out-of-state institutions, including BankAmerica and NationsBank, have penetrated the Texas markets and gathered up more than a third of the state's deposits.[373] This makes interstate branching even less attractive to Texans. While community banks in other states might be persuaded to accede to interstate branching in the hope of being acquired by out-of-state banks for a handsome premium,[374] Texas banks were targets of out-of-state acquisitions in

---

in LEXIS, Bankng Library, Ambank File; Robyn Meredith, *OCC: States May Find No Turning Back on Opting Out of Interstate*, AM. BANKER, Mar. 30, 1995, at 2, *available in* LEXIS, Bankng Library, Ambank File. On the other hand, the decision to join the interstate branching system (whether by early-in election or by operation of law in 1997) is clearly irrevocable. *See* Richard A. Oppel, Jr., *Small Banks Switch Stance on Interstate Branching*, DALLAS MORNING NEWS, Mar. 28, 1995, at D1.

[369] *See* O'Hara, *Governor*, *supra* note 361, at 1.

[370] "The General Assembly finds that it is in the best interests of this state to declare that interstate branching in Colorado is prohibited prior to June 1997." COLO. REV. STAT. § 11-6.4-101 (1995).

[371] TEX. REV. CIV. STAT. ANN. art. 489f § 2 (West 1995).

[372] Dee DePass, *Bankers Join Forces For Study of Interstate Branching Law*, STAR TRIB., Oct. 7, 1994, at D1.

[373] *See supra* note 212 and accompanying text. NationsBank alone controls 20% of the banking assets in Texas. Allison Bell, *NationsBank Fights Battle in Texas for New Bank Bill*, BUS. J.-CHARLOTTE, Mar. 27, 1995, at 7.

[374] This is the reason that community banks may prefer that states adopt an acquisition-only rule, rather than exercising their option under 12 U.S.C. § 36(g)(1)(A) to

the mid-1980s. The handsome premiums, in other words, have already been paid. As a Houston newspaper editorial explained:

> Independent banks in other states mostly favor the [interstate branching] measure, since it requires new banks entering the state to buy an existing bank before they can set up shop. But no such bonanza is likely for Texas bankers looking to sell out to the big banks coming in.
>
> Thanks to the 1980s bust, when failed Texas banks were sold to whomever would take them off our hands, the big boys are already here.[375]

Consequently, Texas chose to opt out of interstate branching, though not without providing for automatic expiration of the opt-out statute as of September 2, 1999.[376]

A handful of other states—including Missouri, Montana, Nebraska, and New Mexico—are also considering removing themselves from the interstate branching regime.[377] Among the arguments being deployed at the state level are those previously advanced at the federal level, as surveyed in Part III. The IBAA, for example, argues that opt-out legislation is needed to avoid overconcentration, deposit siphoning, undue risk to the BIF, and loss of local control over depository institutions. In a leaflet entitled "Why Community Bankers Should Be Interested in Opt-Out Legislation," the IBAA urges its members to support opt-out initiatives in their states:

> [I]nterstate branching will be neither benign to community bankers, as some interstate advocates have suggested; nor will it be a boon for community bankers, as some banking analysts have predicted. Indeed, interstate branching is bad public policy; it provides no benefits for community bankers; and it is little more than special interest legislation to allow our nation's biggest banks to consolidate their empires.
>
> America's banking system is the envy of the world. Let's not dismantle our powerful economic engine that has fostered entrepreneurship and opportunity for all citizens. Interstate branching, if allowed by the states, will exact a devastating price on local communities. Farmers, ranchers, small businesses and consumers all will suffer if interstate branch-

---

authorize entry via *de novo* branching. *See supra* notes 351–352 and accompanying text.

[375] Jim Barlow, *Interstate Banks Aid Competition*, HOUSTON CHRON., Dec. 11, 1994, at 1.

[376] TEX. REV. CIV. STAT. ANN. art. 489f § 3 (West 1995).

[377] *See More States Favor Early Participation in Interstate Branching*, BANKING POL'Y REP., Sept. 18, 1995, at 6.

> ing is allowed to take effect. Act now to protect the financial integrity of your state.[378]

To advance its legislative agenda, the IBAA has drafted a model opt-out statute and established a large fund to which state groups lobbying for opt-out legislation may apply for financial support.[379] Thus, the defenders of the McFadden Act are once again speaking out, this time as part of the state-level response to the federal legislation.

D. *Alternatives to Opting Out*

States that want to protect their community banks without going so far as to opt out have a number of alternatives. One of the consequences of the Riegle-Neal Act's respect for state sovereignty is that the states have the power to put a variety of obstacles in the way of interstate branching. Using the devices discussed above—concentration limits, minimum age laws, and consumer protection laws—a state can create a regulatory environment that is hostile to out-of-state institutions seeking to enter via branching. For example, a state may establish burdensome application procedures, lower the usury ceiling, and develop an antitrust regime that polices mergers and acquisitions in a particularly draconian way.[380] In addition, Congress's explicitly stated intention not to interfere with state tax laws[381] gives the states considerable opportunity to use their tax codes to discourage out-of-state banks from opening branches in-state. Within constitutional limits, a state is free to create a tax climate that disfavors branching into the state.[382] In view of all these options open to the states, the IBAA has emphasized in materials sent to its membership that even if lobbying efforts to opt out fail, the community banks can still demand other forms of protection.[383]

---

[378] Independent Bankers Association of America, Why Community Bankers Should Be Interested in Opt-Out Legislation (undated and unpublished manuscript, on file with author).

[379] *See* Terrence O'Hara, *An Arsenal of Opt Out: $250,000 of Ammo from the IBAA*, AM. BANKER, Feb. 15, 1995, at 6, *available in* LEXIS, Bankng Library, Ambank File.

[380] *See* Bell, *supra* note 373, at 19.

[381] Riegle-Neal Act § 101(b) (amending Bank Holding Company Act § 7(b)).

[382] *See* Bell, *supra* note 373, at 19.

[383] Independent Bankers Association of America, *supra* note 378.

E. *The "Early-in" Option*

The flip side of the state sovereignty principle is that states may also authorize interstate branching early. States are permitted to authorize interstate branching by national banks headquartered within their boundaries in advance of the effective date provided in the Riegle-Neal Act (June 1, 1997).[384] If States *X* and *Y* both authorize interstate branching early, then *X*'s national banks may branch into *Y* and vice-versa, even though the federal statute has not yet become effective. Oregon was the first state, in February 1995, to enact such an "early-in" statute.[385] Now Oregon's national banks and state-chartered Federal Reserve member banks may branch across state lines, just as state-chartered nonmember banks have been authorized to do since 1993.[386] Fourteen other states have also enacted "early-in" legislation, and four states, including Colorado, have passed laws consenting to interstate branching as of the federal target date.[387] In addition, the "early-in" option is under consideration in several other states. Those that are home to large banks with multistate ambitions appear to be the ones most likely to shorten the timetable set by Congress for interstate branching. For example, New York and California—where large, expansion-minded banking organizations like Citicorp, BankAmerica, and First Interstate are based—are among the states likely to grant early approval for interstate branching.[388] These states recognize that it may be perilous to wait for interstate branching to take effect automatically, as it will—for all states that do not opt out—in mid-1997. By that time, banks from the "early-in" states may have already seized the best opportunities for expansion into new markets.[389]

---

[384] "A merger transaction may be approved pursuant to paragraph (1) before June 1, 1997, if the home State of each bank involved in the transaction has in effect, as of the date of the approval of such transaction, a law that . . . expressly permits interstate merger transactions with all out-of-State banks." Riegle-Neal Act § 102(a) (amending Federal Deposit Insurance Act § 44(a)(3)).

[385] 1995 Or. Laws 6.

[386] *See Oregon Governor Signs "Opt-In" Interstate Bank Branching Bill*, DAILY REP. FOR EXECUTIVES, Mar. 8, 1995, at 45, *available in* LEXIS, Exec Library, Drexec File.

[387] *See More States Favor Early Participation in Interstate Branching*, BANKING POL'Y REP., Sept. 18, 1995, at 6.

[388] *See CSBS Outlines State Initiatives to Implement Interstate Branching Law*, BANKING POL'Y REP., Nov. 21, 1994, at 9; Allison Bell, *Big N.C. Banks Face Hurdles to Interstate Growth*, BUS. J.-CHARLOTTE, Jan. 23, 1995, at 19; Terrence O'Hara, *New IBAA Chief Sees a Crossroads Year*, AM. BANKER, Feb. 14, 1995, at 10, *available in* LEXIS, Bankng Library, Ambank File.

[389] *See* Bowers Interview, *supra* note 352, at 5.

F. *Inter-bank Agency Activities*

Some commentators have argued that a state's decision to opt out, "early-in," or simply await 1997 may ultimately have little importance, on account of a provision in the Riegle-Neal Act that allows affiliated banks to act as agents for one another.[390] As discussed in Part II, sibling banks within a common BHC family have traditionally been permitted to furnish certain services and assistance to one another, as long as they do not act in substance like branches of a single bank.[391] The Riegle-Neal Act's agency provision enlarges the range of services that bank subsidiaries may offer one another without being deemed branches: "Any bank subsidiary of a BHC may receive deposits, renew time deposits, close loans, service loans, and receive payments on loans and other obligations as an agent for a depository institution affiliate."[392] Some commentators consider this provision so far-reaching, particularly in its deposit-collection feature, that it renders state opt-out statutes nugatory. In their view, not even a state that has opted out can avoid the agency provision, and will thus experience the equivalent of interstate branching, as nominally independent banks use their new agency powers to behave essentially like branches.[393] This position, however, is exaggerated. Even an agent with substantial powers is still an agent, not a branch. The Joint Explanatory Statement makes clear that an agent can only take deposits for existing accounts, rather than opening new accounts, and can only service loans (i.e., distribute applications, assemble documents, accept loan payments, etc.) without actually making credit decisions or disbursing funds.[394] Thus, despite the alarm raised over this point, the agency provision does not appear to eviscerate the opt-out clause.

---

[390] Riegle-Neal Act § 101(d) (amending Federal Deposit Insurance Act § 18(r), 12 U.S.C. § 1828).

[391] *See supra* note 3 and accompanying text.

[392] Riegle-Neal Act § 101(d) (amending Federal Deposit Insurance Act § 18(r)(1), 12 U.S.C. § 1828).

[393] *See* Olaf de Senerpont Domis, *Loophole in Interstate Law Offers Benefits of Branching in '95, Nearly 2 Years Early*, AM. BANKER, Nov. 16, 1994, at 3, *available in* LEXIS, Bankng Library, Ambank File; *States Face Deadline in Interstate Branching Law That Limits Opt-Out Choice*, BNA'S BANKING REP., Nov. 14, 1994, at 694.

[394] *See* Joint Explanatory Statement, *supra* note 338, at H6637.

G. *The Statute's Responses to Opponents' Concerns*

Although the preservation of states' rights is a major theme of the new legislation, there is more to the Riegle-Neal Act than federalism. The statute is also noteworthy for the way it responds to the concerns raised by opponents of interstate banking and branching. A number of its provisions are designed to control for precisely the problems that critics predict will result from deregulation. Safety and soundness concerns, for example, are addressed by the requirement that banks branching across state lines be both adequately managed and adequately capitalized.[395] Similarly, the concentration limits are aimed at preventing banks from acquiring excessive market power. Placing ten and thirty percent caps on banks' national and state market shares, respectively, is a sensible solution to the problem because of its directness: Congress has replaced a blanket anti-branching rule with a narrowly tailored prohibition.[396] Depending on the perceived need for size control, the states may wish to ratchet this thirty percent figure up or down, as permitted by the Riegle-Neal Act.[397] Although deposit caps are nothing new—about one-third of the states already have caps, ranging from ten to thirty percent—they are expected to become more important as a means of regulating bank expansion.[398] Rather than wrangling over whether to opt out, bankers and lawmakers may prefer to focus the debate on where to set the state concentration limit.[399]

The Riegle-Neal Act also contains a number of checks on a branch's ability to drain funds out of its community. First, the statute specifies that all interstate merger and branching transactions are subject to the CRA.[400] CRA evaluations are to be made of a bank's performance not only overall (i.e., in all its markets nationwide, irrespective of state boundaries), but also in each state and metropolitan area where it maintains branches.[401] Admittedly, this procedure is not perfectly refined, in that it does

---

[395] Riegle-Neal Act § 102(a) (amending Federal Deposit Insurance Act § 44(b)(4)).

[396] *See* Miller, *supra* note 74, at 225.

[397] Riegle-Neal Act § 101(a) (amending Bank Holding Company Act § 3(d)(2)(C)).

[398] For an extensive discussion of concentration limits, *see* John C. Coates IV & David S. Neill, *New Interstate Moves Still Face Minefield of Deposit Cap Statutes*, BANKING POL'Y REP., Aug. 15, 1994, at 23.

[399] *See, e.g.*, Terrence O'Hara, *Washington Banks Seek Tighter Lid on Deposit Share*, AM. BANKER, Jan. 5, 1995, at 7, *available in* LEXIS, Bankng Library, Ambank File.

[400] Riegle-Neal Act § 102(a) (amending Federal Deposit Insurance Act § 44(b)(3)).

[401] *Id.* § 110(a) (amending Community Reinvestment Act, 12 U.S.C. § 2906(d)(1)).

not require CRA reports to be filed on a branch-by-branch basis. There is a certain injustice to the requirement that an independent bank must undergo a full CRA evaluation, while its competitor next door—a branch of an out-of-state bank—may succeed in escaping examination altogether. As an IBAA representative told Congress: "[O]ur members—which are fully examined for CRA compliance—will often compete with branches of larger regional or nationwide banks that may never be visited by a CRA examiner . . . .[N]o branch of an interstate bank will face the same level of CRA scrutiny as a community bank."[402] Chiefly for reasons of administrability, the Riegle-Neal Act does not require branch-specific CRA assessments. The state- and city-specific examinations, however, can be expected to prevent multistate banks from adopting lending practices that favor one market over another.

In addition to CRA monitoring, the Act provides two further safeguards against deposit siphoning. An out-of-state bank seeking to close a branch located in a low- or moderate-income area must provide advance notice to the community, and community leaders may require that bank officials meet with them to discuss alternative means of obtaining credit.[403] Further, out-of-state branches are prohibited from serving as mere deposit production offices for the main bank.[404] They must show activity on both sides of the balance sheet: just as they borrow in the form of collecting deposits, so too must they lend. The statute does not require that they lend in proportion to their borrowing, but their level of community reinvestment must be at least half of the average level prevailing in the host state. According to the statute, regulatory authorities may require closure of an out-of-state branch

> if the appropriate Federal banking agency for the out-of-State bank determines that the bank's level of lending in the host State relative to the deposits from the host State . . . is less than half the average of total loans in the host State relative

---

[402] *Hearings Before the Subcomm. on Financial Institutions and Consumer Credit of the House Comm. on Banking and Financial Services*, 104th Cong., 1st Sess. (1995) (testimony of Tony Abbate, on behalf of the Independent Bankers Association of America).

[403] Riegle-Neal Act § 106 (amending Federal Deposit Insurance Act, 12 U.S.C. § 1831r-1(d)).

[404] A bank is prohibited from "using any authority to engage in interstate branching pursuant to this title, or any amendment made by this title to any other provision of law, primarily for the purpose of deposit production." *Id.* § 109(a).

> to total deposits from the host State . . . for all banks the home State of which is such State . . . .[405]

In this way, the Riegle-Neal Act ensures that banks will properly support local interests when they branch across state lines.

Thus, not only does the new statute make needed reforms to the banking system, but it is also sensitive to the concerns of its detractors. The Riegle-Neal Act is a model of responsible, not reckless, deregulation. Congress has provided a means for developing a seamless, coast-to-coast banking system, but at the same time it has recognized that such a system must initially be experimental. The statute leaves ample room for dissent and even backtracking at the state level. Whereas it clearly seeks to debunk the myth that the state is the largest banking area that can be authorized consistent with safe and sound principles, the Riegle-Neal Act nevertheless respects the deeply held American conviction that the states are often better positioned than the federal government to make regulatory judgments. On account of the discretion reserved to the states, the Riegle-Neal Act does not modernize the bank regulatory structure as quickly or as completely as some bankers and commentators would like. Until the states have fully responded to the Riegle-Neal Act—by opting out or joining in early, setting conditions on entry by out-of-state banks, updating their tax codes, and addressing the many other issues raised by the statute—the scope of interstate banking and branching will remain unknown.

V. Toward the Next Century

The Riegle-Neal Act is landmark legislation in the bank regulatory field, if only because it requires that two other landmarks be razed. The American banking landscape will look significantly different in the absence of the McFadden Act and the Douglas Amendment. Although the repeal of these statutes is unlikely to set off a firestorm of merger and acquisition activity,[406] further consolidation of the industry has begun and will doubtless con-

---

[405] *Id.* § 109(c)(1). Suppose, for example, that a bank headquartered in State *A* established a branch in State *B*, where it collected $20 in deposits and extended $5 of credit. Collectively, the State *B* banks gathered $200 in deposits from State *B*, and made $150 worth of loans there. The branch would be subject to closure under the statutory test, because 0.25 (i.e., 5 divided by 20) is less than half of 0.75 (i.e., 150 divided by 200). The branch would have to increase its lending or reduce its deposits.

[406] *See* Streeter & Cocheo, *supra* note 302, at 45.

tinue.[407] Large banks are the most likely acquirers, small and mid-sized banks the most likely targets. Healthy banks are more likely to be both targets and acquirers. Banks will tend to expand into states and regions with vibrant economies, while avoiding depressed areas.[408] The Riegle-Neal Act will encourage the growth of large, geographically diversified institutions, and it will become more common for banks to pursue multistate and multiregional strategies.[409] Small banks, however, will surely remain a viable part of the banking industry. Federal Reserve Governor John LaWare predicts that the United States will always have 7000–8000 banks, the vast majority of them independent.[410] Indeed, the consensus is that, even after further industry consolidation, the United States will still be home to at least 5000 banks at the turn of the century.[411]

There may also emerge a handful of genuinely nationwide banks, having coast-to-coast branching networks. Such institutions, serving customers in precisely the same way throughout the country, would represent a dramatic change from the status quo. These nationwide banks would be able to achieve a level of standardization in their products and services heretofore unknown in the American banking industry.[412] As a BankAmerica executive explained: "What we want to do is provide a single image to our customers across geographic lines."[413] In the post-

---

[407] *See* Christine Dugas & David Henry, *Big Year for Bank Deals: First Interstate Faces $10 Billion Unsolicited Bid*, USA TODAY, Oct. 19, 1995, at B1; Uri Berliner, *Corporate Wedding Bells Become a Din*, SAN DIEGO UNION-TRIB., Sept. 16, 1995, at A1.

[408] *See* GAO 1993, *supra* note 10, at 68–69 (citing John Shoenhair & Kenneth Spong, INTERSTATE BANK EXPANSION: A COMPARISON ACROSS INDIVIDUAL STATES, STUDY FOR THE FEDERAL RESERVE BANK OF KANSAS CITY (1990)); Savage, *supra* note 51, at 1085.

[409] *See* GAO 1993, *supra* note 10, at 47, 56; MCCOY ET AL., *supra* note 51, at 156.

[410] *Regulatory Relief, Expanded Powers on Horizon for Banks, LaWare Tells IBAA*, 64 BNA'S BANKING REP., 361, 362 (1995).

[411] *See, e.g.*, Bill Atkinson, *Keycorp President Says Small Banks Need Less Red Tape than Big Ones*, AM. BANKER, Mar. 10, 1995, at 7, *available in* LEXIS, Bankng Library, Ambank File, (interview with Robert W. Gillespie, president, Key Corp); MCCOY ET AL., *supra* note 51, at 55.

[412] One banking consultant emphasized this point: "The primary impact on bank operations would be an opportunity and challenge to standardize delivery systems across the country such that someone walking in a branch anywhere will see the same set of services delivered in the same way." Deidre Sullivan, *How Will the Operations of Large Banks Be Changed If Interstate Branch Banking Laws Are Loosened?*, AM. BANKER, Sept. 26, 1994, at 10A, *available in* LEXIS, Bankng Library, Ambank File, (quoting Lawrence A. Willis, executive vice president, First Manhattan Consulting Group).

[413] Hoffman, *supra* note 157, at 53 (quoting Jim Burke, senior vice president, BankAmerica).

McFadden world, this would be more than merely the image of a family of commonly owned but separately chartered banks. It would be the image of a single, unified institution.

A. *The European Experience*

It would be careless, however, to suggest that the removal of geographic barriers will inevitably result in the establishment of Canadian-style nationwide banks in the United States. To be sure, the Riegle-Neal Act makes such nationwide institutions theoretically possible. However, the EU's Second Banking Directive made pan-European banks possible as well, and little has come of it. For three years the EU has had a "single banking license," pursuant to which a bank chartered by one EC member state may branch freely into the others. In practice, however, banks have not widely taken advantage of this privilege.[414] The banking market in Europe remains extremely fragmented; many banks are national, but few are multinational. In a 1994 publication, the European Commission observed:

> The European retail banking market is still very much made up of the national markets. In recent years there has been a spate of mergers, acquisitions, alliances and cooperations between banks, but many of these deals involved only domestic banks. Many national champions were created but few international ones.[415]

The only multinational banks in Europe are Crédit Lyonnais of France, Deutsche Bank of Germany, and Barclays of the United Kingdom. Crédit Lyonnais has 900 banking facilities outside France, mostly acquired in the late 1980s. It plans eventually to merge them into a single pan-European branch network. Although Crédit Lyonnais has recently suffered severe financial difficulties, its European banks have performed relatively well.[416]

---

[414] *See* Golembe, *supra* note 132, at 6 ("Great things were expected from the creation of a single banking market in Europe . . . . But . . . nothing much has happened."); *Being All You Want to Be, International Banking Survey*, ECONOMIST, Apr. 30–May 6, 1994, at 32, 32 (separately paginated survey at 62); *EU: A Case of Eurosclerosis?*, MORTGAGE FIN. GAZETTE, Apr. 6, 1994, available in LEXIS, Bankng Library, Morfin File.

[415] EUROPEAN COMMISSION, PANORAMA OF EU INDUSTRY 94, at 23-16 (1994) [hereinafter PANORAMA].

[416] *See* Alice Rawsthorn & David Buchan, *Bank Takes Long March to Profit*, FIN. TIMES (London), Apr. 6, 1994, at 24.

Similarly, Deutsche Bank has made major acquisitions in Spain and Italy and clearly has pan-European ambitions.[417] Barclays has acquired hundreds of facilities in France and Spain, and along with Crédit Lyonnais and Deutsche Bank, it is one of the few banks whose name has become a household word throughout much of Europe.[418] These accomplishments are somewhat diminished, however, by the fact that two banks from across the Atlantic have achieved the same result. Citibank and Chase Manhattan (newly merged with Chemical Bank) are as widespread and well known in Europe as any native European bank. Citibank alone has more than 800 offices across Europe.[419] It is ironic that despite the single banking license being available only to native European banks, U.S. banks have had as much success expanding across the continent as their European counterparts.

The Second Banking Directive cannot, therefore, be said to have stimulated in practice the kind of behavior expected in theory. Concededly, unlike U.S. institutions, expansion-minded European banks face language and cultural barriers that make cross-border branching particularly challenging. Nevertheless, the European experience is instructive. Merely giving banks legal authorization to branch long-distance does not necessarily help them overcome the difficulties of managing far-flung enterprises and attracting loyal customers away from local institutions.[420] Rather than branching across borders, European banks have tended to prefer strategic alliances, joint ventures, and other types of cooperation agreements with banks in neighboring countries.[421] This may suggest something about how the U.S. banking industry will react to the Riegle-Neal Act. Perhaps banks will still consider loose affiliation with other banks a more efficient

---

[417] *See* Kevin Muehring, *Kopper Takes the Helm at Deutsche Bank*, INSTITUTIONAL INVESTOR, July 1992, at 183; Ferdinand Protzman, *Deutsche Bank Subsidiary Acquires Banco de Madrid*, N.Y. TIMES, Mar. 6, 1993, at 37; John Gapper & Andrew Fisher, *Hopeful Signs of a Reversal of Fortune: Hilmar Kopper's Plans to Rebuild Deutsche Bank's Reputation*, FIN. TIMES (London), Nov. 7, 1994, at 19.

[418] *See Barclays Pursues Europe*, FIN. TIMES (London), Jan. 16, 1992, at 42; David Buchan, *Barclays Expands in France*, FIN. TIMES (London), Oct. 23/24, 1993, at 10.

[419] *See* Wilmarth, *supra* note 9, at 1068; Patrick Oster, *Citibank Branching Out in Europe*, NEWSDAY, June 16, 1993, at 49, *available in* LEXIS, News Library, Newsdy File; Robert Selwitz, *Window on Eastern Europe: Citicorp Asks East Germans to Empty Their Socks*, ABA BANKING J., July 1992, at 7.

[420] Wilmarth, *supra* note 9, at 1067, 1068.

[421] *Id.* at 1067; PANORAMA, *supra* note 415, at 23-13 to 23-14.

and effective interstate strategy than making the substantial investment required to establish actual branches.

B. *The Role of Technology*

Visions of grand branching networks stretching across the United States also conflict with an important recent trend in consumer behavior. Many people, particularly those under age thirty-five, would rather bank electronically than visit a branch. To serve this segment of the market, banks are better off investing in technology than in branch facilities. Young consumers want to bank via ATM, telephone, and personal computer.[422] High-tech banking is also preferable from the bank's perspective. The traditional brick-and-mortar delivery system is expensive and inefficient. A live teller transaction costs the bank on average $1.07, compared with 35 cents for a phone call and 27 cents for an ATM visit.[423] Ironically, just as the branching laws are being liberalized, banks have an incentive to phase out their branching operations. Although Congressman Neal implied that his proposal would encourage establishment of branches,[424] it is widely forecast that thousands of branches will be closed by the end of the decade.[425] Simultaneously, however, banks must cater to an older generation of consumers who demand face-to-face contact with branch staff, thus ensuring the survival of the traditional branch. In the years to come, banks will be required to excel in both branch and non-branch service.[426]

C. *The Fate of the BHC Structure*

To the extent that the Riegle-Neal Act stimulates long-distance branching, the banks best positioned to take advantage of the new statutory scheme are those that have been most successful at expanding via BHC: the superregionals. Leading candidates

---

[422] *See* Gibson, *supra* note 101, at 155; John Russell, *Trends in Retail Bank Distribution*, *in* McCOY ET AL., *supra* note 51, at 234–35.

[423] Gibson, *supra* note 101, at 155.

[424] 140 CONG. REC. H1861 (daily ed. Mar. 22, 1994) (statement of Representative Neal).

[425] *See* McCOY ET AL., *supra* note 51, at 227; Gibson, *supra* note 101, at 155.

[426] *See* McCOY ET AL., *supra* note 51, at 223, 228–29; Gibson *supra* note 101, at 155; *For Our Next Trick, International Banking Survey*, ECONOMIST, Apr. 30–May 6, 1994, at 25, 28–31 (separately paginated survey at 62).

for multiregional expansion include NationsBank, BancOne, BankAmerica, KeyCorp, First Union, PNC, Norwest, and Fleet. These organizations, along with a few other superregionals, are by most accounts the stars of today's banking industry.[427] The superregional formula has yielded levels of market capitalization[428] significantly higher than those of the money-center institutions. The money-center banks are not likely to be major players in the latest wave of expansion and acquisitions.[429] Instead, the superregionals can be expected to lead the way into the next century, as they move from a regional to a multiregional philosophy.

Some BHCs will immediately take advantage of the opportunity to merge their subsidiary banks into a single multistate institution. NationsBank, having lobbied extensively for the new legislation, will undoubtedly consolidate its large empire in an effort to realize cost savings. Other BHCs, however, have expressed an intention to retain their existing bank subsidiary structures. BancOne, for example, prefers a decentralized system in which distinct legal entities are loosely united under a common corporate umbrella. On this view, separate boards of directors and management actually add value—enough to offset the additional overhead costs—because of their familiarity with the community and their ability to attract customers who prefer local banks. BancOne also believes that bankers drawn from the community, especially when given sufficient autonomy, have more incentive to achieve results than personnel simply transplanted from the home office. Someone whose title is "president" will feel she has a greater personal stake in the success of the enterprise than if she is called "bank manager."[430] NationsBank would respond to this argument by asserting that it, too, prides itself on sensitivity to the community. McColl has stressed this factor:

---

[427] *See* Danielson, *supra* note 212, at 11; McCoy ET AL., *supra* note 51, at 139, 146–48, 154.

[428] Market capitalization is the "value of a corporation as determined by the market price of its issued and outstanding stock. It is calculated by multiplying the number of outstanding shares by the current price of a share." JOHN DOWNS, DICTIONARY OF FINANCE AND INVESTMENT TERMS 246 (Jordan E. Goodman ed., 3d ed. 1991).

[429] *See* McCoy ET AL., *supra* note 51, at 123, 125, 141, 154; GAO 1993, *supra* note 10, at 42; Steve Lohr, *Recasting the Big Banks: Weakened Giants, in Humbling Mergers, Are Fighting to Regain Their Dominance*, N.Y. TIMES, July 17, 1991, at A1, D6.

[430] McCoy ET AL., *supra* note 51, at 58, 338; Wilmarth, *supra* note 9, at 1006; *Debate: What Is the Future of Banking?*, HARV. BUS. REV., July–Aug. 1991, at 144, 151 (remarks by Robert E. Litan).

"We empower the local manager to make local decisions, and we run them as if they are community banks."[431] Thus, there is agreement over the importance of keeping close to the community, but disagreement over which legal form should be used in implementing this strategy.[432]

Given the BancOne point of view, the BHC device does not appear to be destined for extinction. Many scholars and practitioners have argued that BHCs would not exist absent the need to circumvent the McFadden Act. BHCs were developed as a way of crossing state lines indirectly so as not to violate the prohibition on interstate branching. Now that this stratagem is no longer necessary, perhaps the BHC format may be discarded. Attorney Melanie Fein makes the point bluntly: "Bank holding companies have no inherent use in the banking system."[433] BancOne, however, clearly believes that the BHC is a useful vehicle, and others note that BHCs provide advantages, such as allowing banking organizations to bypass restrictions on bank powers and activities.[434] Pursuant to 12 U.S.C. § 24 (Seventh), a national bank may exercise only those powers "necessary to carry on the business of banking," such as receiving deposits, lending money, and issuing notes.[435] Banks are thus forbidden to engage in commerce. Section 16 of the Glass-Steagall Act further prohibits banks from underwriting, selling, or dealing in securities.[436] Using the BHC device, however, a banking organization can enlarge its field of operations. In addition to banks, BHCs are permitted to own companies whose activities are "so closely related to banking or managing or controlling banks as to be a proper incident thereto . . . ."[437] Hence, a BHC's nonbank subsidiaries may engage in activities such as providing data processing services, courier services for banking documents,

---

[431] *Mar. 1991 Hearing*, *supra* note 147, at 577 (testimony of Hugh McColl).

[432] This strategy has been termed "supercommunity banking." In a short essay, Anat Bird sets out the main components of this philosophy, which allows mid-sized banks to "outlocal the nationals and outnational the locals." Bird concludes that this is the optimal strategy: "The [supercommunity bank] takes the community banking approach to the customer from the small bank and combines it with some measure of cost efficiencies, which are created by economies of scale and the breadth of product offering, from the large banks. That combination is a winning strategy for banks into the next century." Anat Bird, *Supercommunity Banking*, *in* McCoy ET AL., *supra* note 51, at 252–54.

[433] Fein, *supra* note 324, at 15.

[434] *See* Savage, *supra* note 51, at 1080; GAO 1993, *supra* note 10, at 83.

[435] 12 U.S.C. § 24 (Seventh) (1994).

[436] *See supra* note 3.

[437] 12 U.S.C. § 1843(c)(8) (1994).

securities brokerage services, and tax planning and preparation services, as well as giving investment and management consulting advice.[438] With or without the McFadden Act, the BHC format will still be needed to enable banking organizations to conduct such operations.

On the other hand, two distinguished practitioners have recently argued that the "closely related to banking" test has been interpreted so narrowly that BHCs ultimately permit few additional activities. Melanie Fein takes an extreme position: "There are virtually no activities in which a bank holding company may engage that a national bank may not conduct."[439] She advocates moving all of a BHC's nonbank operations to the bank subsidiary itself, or to a subsidiary of the bank, and jettisoning the BHC apparatus.[440] Adopting essentially the same position, H. Rodgin Cohen focuses on the particularly thorny problem of securities powers, where the consensus has always been that BHCs clearly offer advantages over banks.[441] Based on a subtle analysis of Sections 16 and 20 of the Glass-Steagall Act, Cohen concludes that a state nonmember bank, and perhaps even a national bank, might be able to hold a securities subsidiary directly, rather than having a BHC hold it as an affiliate of the bank.[442] The BHC would then be superfluous. However, even if this rather tenuous argument were theoretically correct, banking organizations are likely in practice to prefer the status quo. Cohen concedes:

---

[438] Federal Reserve Board Regulation Y, 12 C.F.R. § 225.25 (1995).

[439] Fein, *supra* note 324, at 15.

[440] *Id.*

[441] Pursuant to 12 U.S.C. § 24 (Seventh) (1994), banks may not underwrite securities. A nonbank subsidiary of a BHC, however, may engage in a certain amount of underwriting activity. Section 20 of the Glass-Steagall Act prohibits affiliation between banks and firms "engaged principally in the issue, flotation, underwriting, public sale, or distribution . . . of stocks, bonds, debentures, notes or other securities . . . ." 12 U.S.C. § 377 (1994). A firm that derives no more than 10% of its gross revenues from underwriting is considered not to be "engaged principally" in underwriting. Securities Indus. Ass'n v. Board of Governors of the Fed. Reserve System, 839 F.2d 47 (2d. Cir.), *cert. denied*, 486 U.S. 1059 (1988).

[442] The argument is that Section 16 of the Glass-Steagall Act only places securities restrictions on national banks. 12 U.S.C. § 24 (Seventh) (1994). Section 5(c) in turn applies those restrictions to state member banks, but not to state nonmember banks. 12 U.S.C. § 335 (1994). Hence, a state nonmember bank might be able to hold a securities subsidiary directly. As for national banks, the argument—more implausible still—is that perhaps 12 U.S.C. § 24 (Seventh) (1994) only prohibits national banks from buying and selling securities for the purpose of investment or speculation. This prohibition may not necessarily contemplate holding stock in a subsidiary. Consequently, the subsidiary of a national bank might be able to engage in a broader range of activities than the bank itself. Cohen, *supra* note 79, at 9.

> To some extent, the holding company structure may still afford greater flexibility or powers, or at least there may be some question as to whether it does. Of at least equal importance, banking organizations will be reluctant to undertake a major corporate restructuring unless it can identify a clear benefit.[443]

Thus, the Riegle-Neal Act does not appear to spell the death of the BHC. The more sensible conclusion is that of the GAO, which indicated in its 1993 report that "it is likely that both forms of organization will survive under interstate branching."[444] Some banking companies can be expected to experiment with the new single-bank format, while others retain the BHC structure.

D. *The Fate of the Dual Banking System*

Another feature of the present system that is likely to endure is the dual banking system and the competition between state and federal charters. Many believe that multistate banks will tend to prefer the national charter, in part because they would then be subject to only one regulatory authority.[445] The states, however, are aware of the risk that they may lose big banks to the national charter. In conjunction with the Conference of State Bank Supervisors (CSBS), many states are now reexamining their regulatory systems and developing ways to make their charters more attractive to multistate banking organizations.[446] A state charter may be desirable for several reasons: state supervision is less costly to banks than OCC supervision, state banks are often afforded a wider range of powers than national banks, and the state regulatory environment may be more accommodating. In addition, the advent of interstate branching will provide the states with an incentive to improve interstate cooperation in bank examination and supervision, to ensure that the regulatory bur-

---

[443] Cohen, *supra* note 79, at 9.

[444] GAO 1993, *supra* note 10, at 82.

[445] *See, e.g.*, Savage, *supra* note 51, at 1082; William M. Isaac, *Comment: Miracle of the Marketplace Applies to Regulators, Too*, AM. BANKER, Dec. 15, 1994, at 5, *available in* LEXIS, Bankng Library, Ambank File. *But see* Geoffrey P. Miller, *The Future of the Dual Banking System*, 53 BROOK. L. REV. 1 (1987) (arguing that the elimination of geographic restrictions will render national charters unnecessary).

[446] *See* James B. Watt, *Actions by States Will Shape Transition to Interstate Branching*, BANKING POL'Y REP., Aug. 15, 1994, at 27, 27; *State Regulators Facing Deadline in Dealing with Interstate Branching*, DAILY REP. FOR EXECUTIVES, Nov. 4, 1994, at 212, 212. *available in* LEXIS, Exec Library, Drexec File.

den associated with a state charter will be no greater than for a national charter.[447] The CSBS is optimistic that the state charter will continue to have appeal:

> The state system, building on its intrastate strength, will configure itself to present a viable alternative to a national bank charter, for those banks choosing to operate on an interstate basis . . . . Not only will state charters be a reasonable competitive alternative for interstate banks, state chartering may well become the charter of choice.[448]

Thus, it appears that the Riegle-Neal Act will not destroy the dual banking system, but reenergize it.[449] As the President of the CSBS observed, the new law represents simply "another stage in the evolution of the dual banking system,"[450] rather than a sharp break with the American tradition of competition in chartering.

CONCLUSION

This sketch of the world created by the Riegle-Neal Act suggests that the American banking system in the years to come will be characterized by a mix of the old and the new. The statute does not take a wrecking ball to the banking industry, or to traditional notions of the close relationship between banks and their communities. Instead, it builds on the present structure by authorizing interstate branching as an alternative to existing strategies. It permits the establishment of large, multiregional branching networks, but at the same time leaves undisturbed such familiar institutions as the community bank and the BHC family of subsidiary banks. Judging from the profusion of states' rights provisions, as well as the clause that sets the effective date for interstate branching almost three years in the future, it is clear that Congress did not want to force change too abruptly. The Riegle-Neal Act is the product of a legislative body committed to making steady, even progress toward a modernized

---

[447] *See* Isaac, *supra* note 445, at 5.

[448] Conference of State Bank Supervisors, A Vision for the Development of Interstate Branching Regulation and Supervision of State Chartered Banks 5 (Sept. 29, 1994) (unpublished manuscript on file with author).

[449] John Rippey of the Bankers Roundtable points out: "What's healthy for the banking industry is that this has revitalized the dual banking system." *State Regulators Facing Deadline in Dealing with Interstate Branching in* Conference of State Bank Supervisors, *supra* note 448, at 212.

[450] Watt, *supra* note 446, at 27.

banking system. Hugh McColl is right to describe the Riegle-Neal Act as "easily the most significant banking legislation in more than a half-century," and he may even have a point when he predicts that it "will effect [sic] virtually every bank customer in America."[451] At the same time, it is essential to see that the statute does not attempt to exert any sudden or overwhelming influence. It is simply part of an ongoing process.

[451] *The Future of Banking: The Chairman and CEO of NationsBank Gives a Glimpse of the Future with Interstate Banking*, ST. LOUIS COM., Jan. 1995, at 36 (excerpts from speech by Hugh McColl). It should be noted that McColl, a central figure in getting the legislation passed, had personal cause for celebration. According to a proxy statement filed with the Securities and Exchange Commission by NationsBank, McColl's compensation for 1994 ($13.9 million) was five times what it had been the previous year, in part as a reward for his lobbying success in connection with the Riegle-Neal Act. *NationsBank Chief Got Fivefold Raise*, ATLANTA J. & CONST., Mar. 28, 1995, at D3.

NOTE

WHEN THE RULES DON'T FIT THE GAME: APPLICATION OF THE UNIFORM CHILD CUSTODY JURISDICTION ACT AND THE PARENTAL KIDNAPING PREVENTION ACT TO INTERSTATE ADOPTION PROCEEDINGS

GREG WALLER*

In light of increasing interstate custody disputes between adoptive and biological parents, courts must often engage in complex jurisdictional battles. In an effort to resolve these conflicts, courts have turned to the provisions of the Uniform Child Custody Jurisdiction Act (UCCJA) and the Federal Parental Kidnaping Prevention Act (PKPA)—statutes originally designed to address interstate abductions of children. In this Note, Mr. Waller describes how the use of these statutes in interstate custody disputes produces unintended consequences. In addition, the author illustrates how the often inconsistent judicial application of these procedural rules may have substantive law effects. Mr. Waller proposes possible legislative and judicial reforms to ensure that the application of these statutes will reflect their ultimate goal, namely, the best interests of the child.

> We are in a whole new ball game, where the old rules no longer apply and the new rules haven't been written. Small wonder that custody disputes have been called "potential interstate nightmares."[1]

On February 13, 1995, the United States Supreme Court denied an application to stay the custody order of the Illinois Supreme Court in the case of *O'Connell v. Kirchner*,[2] thus bringing to an end the latest heavily publicized courtroom saga pitting natural parents against adoptive parents in a child custody struggle. In this case, popularly known as the "Baby Richard" case, the Illinois Supreme Court awarded Otakar Kirchner custody of his biological son, who had lived almost from birth with a prospective adoptive family. One month before the child's birth,

---

* Judicial Law Clerk, Judge Emilio M. Garza, United States Court of Appeals, Fifth Circuit; B.A., Vanderbilt University, 1992; J.D., Harvard Law School, 1995. The author wishes to thank Professor David Westfall for his advice and assistance.

[1] *Parental Kidnaping Act of 1979, S. 105: Joint Hearing on S. 105 Before the Subcomm. on Criminal Justice of the Senate Comm. on the Judiciary and the Subcomm. on Child and Human Development of the Senate Comm. on Labor and Human Resources*, 96th Cong., 2d Sess. 2 (1980) [hereinafter *PKA Hearings*] (opening statement of Sen. Mathias).

[2] 115 S. Ct. 1084 (1995).

Mr. Kirchner had ended his relationship with Daniela Janikova in Illinois and returned to his native Czechoslovakia. After the birth of the child in March of 1991, Daniela consented to an adoption, refusing to identify the father. When Otakar later contacted her from Czechoslovakia, Daniela told him that the baby had died. In May of 1991, the couple reconciled. Upon learning that his son was alive, Otakar challenged the adoption.[3] After four years of court battles, the Illinois Supreme Court ordered the only parents that Baby Richard had ever known to surrender custody to the natural father.[4]

Custody disputes in adoption cases often provide heart-wrenching stories. Natural and adoptive parents frequently wage extensive court battles, each party asserting the best interests of the child as his or her primary motivation.[5] Years often elapse before all custody issues are resolved. Caught in the middle of the case is a child, whose sense of stability and continuity, so essential for healthy development,[6] is undermined by a pervading atmosphere of uncertainty during the pendency of the litigation.[7] At the conclusion of the litigation, regardless of the outcome, the child may be uprooted from familiar surroundings during the most formative years of childhood, placed in a home the child has never visited, given a name the child has never heard, and introduced to parents the child has never seen.[8]

Resolution of the legal issues in contested adoptions is even more problematic when the natural parents and the adoptive parents reside in different states. Custody battles between fami-

---

[3] *See* David Bailey, *Court Asked to Make "Tough" Choice in Fight Over Adopted Tot*, CHI. DAILY L. BULL., Sept. 7, 1993, at 1.

[4] *In re* Petition of Otakar Kirchner, 649 N.E.2d 324 (Ill. 1995).

[5] Sadly, courts can rarely make the following observation: "Indeed, the one heroic moment in this litigation has been [the biological mother's] consequent admission . . . that if it were truly in the child's best interests to remain where he is, she would not oppose the adoption." *In re* B.B.R., 566 A.2d 1032, 1045 (D.C. 1989) (Farrell, J., concurring).

[6] *See* JOSEPH GOLDSTEIN ET AL., BEYOND THE BEST INTERESTS OF THE CHILD 31–39 (1979). This book, co-authored by Goldstein, Anna Freud, and Albert J. Solnit, has been extremely influential in the development of child custody law. *See infra* notes 77–79 and accompanying text.

[7] For a judicial description of the negative effects of prolonged custody proceedings, *see, e.g.*, Ramey v. Thomas, 382 So. 2d 78, 80 (Fla. Dist. Ct. App. 1980) ("This case is a classic example of parties warring over a child to such an extent that the primary issue—the welfare and best interest of the child—got lost in the gunsmoke."); *see also PKA Hearings*, *supra* note 1, at 105–06 (testimony of Dr. Jeanette I. Minkoff, probation-family services coordinator) (outlining traumatic effects on children of extended divorce litigation).

[8] *See infra* notes 149–157 and accompanying text.

lies often turn into jurisdictional disputes between courts.[9] In 1993, the legal struggle over "Baby Jessica" caught the public's attention when facts very similar to those involved in the "Baby Richard" case were further complicated because the adoptive parents resided in Michigan while the biological parents lived in Iowa.[10] In cases where the adoptive and biological parents reside in different states, the potential exists for the highest courts of the respective states to issue conflicting orders in the same case, even when applying identical legal standards.

In search of a mechanism for deciding jurisdictional disputes in interstate adoption cases, many courts have invoked the provisions of the Uniform Child Custody Jurisdiction Act (UCCJA)[11] and the federal Parental Kidnaping Prevention Act (PKPA).[12] These statutes were originally enacted in response to interstate abductions of children by noncustodial parents.[13] Leg-

---

[9] *See* HOMER H. CLARK, JR., THE LAW OF DOMESTIC RELATIONS IN THE UNITED STATES § 20.3, at 872 (2d ed., student ed. 1988) ("Subject matter jurisdiction over adoption can fairly be characterized as both confused and uncertain, largely because the statutes are unclear and because there is little reasoned discussion in the cases.")

The Restatement deems subject matter jurisdiction proper in adoption cases if (1) the state is the domicile of either the adopted child or adoptive parent, and (2) the adoptive parent and either the child or custodial party are subject to personal jurisdiction in the state. RESTATEMENT (SECOND) OF CONFLICT OF LAWS § 78 (1969). However, the Restatement position has been described as "not very solidly based on the case law." Clark, *supra*, § 20.3, at 872.

[10] *See In re* Clausen, 502 N.W.2d 649 (Mich. 1993). For an extensive discussion of the "Baby Jessica" case—especially significant because it was written by counsel for the biological parents—see Marian L. Faupel, *The "Baby Jessica Case" and the Claimed Conflict Between Children's and Parents' Rights*, 40 WAYNE L. REV. 285 (1994).

[11] UNIF. CHILD CUSTODY JURISDICTION ACT, 9 U.L.A. 115 (1988).

[12] Parental Kidnaping Prevention Act of 1980, Pub. L. No. 96-611, §§ 6–8, 94 Stat. 3566, 3568 (full faith and credit provisions codified at 28 U.S.C. § 1738A (1988)). The full faith and credit provisions of the Act are consistently referred to in the legislative history as the "heart of the plan." *See, e.g., Parental Kidnaping: Hearing on H.R. 1290 Before the Subcomm. on Crime of the House Comm. on the Judiciary*, 96th Cong., 2d Sess. § 82 (1980) [hereinafter *Parental Kidnaping Hearings*] (statement of Mark M. Richard, Deputy Assistant Attorney General, Criminal Division, U.S. Department of Justice); *PKA Hearings*, *supra* note 1, at 135 (testimony of Russell M. Coombs, associate professor, Rutgers University School of Law).

[13] *See Parental Kidnaping Hearings*, *supra* note 12, at 1 (opening statement of Rep. Conyers) ("The subcommittee takes note of the fact that some estimated 25,000 children are abducted each year by the losing parent in violation of child custody and visitation court orders in the aftermath of divorce proceedings."); *Parental Kidnaping Act of 1979, S. 105: Addendum to Joint Hearing on S. 105 Before the Subcomm. on Criminal Justice of the Senate Comm. on the Judiciary and the Subcomm. on Child and Human Development of the Senate Comm. on Labor and Human Resources*, 96th Cong., 2d Sess. 138 (1980) [hereinafter *Addendum to PKA Hearings*] (additional statement of Sen. Wallop) (defining scope of the Act to extend "its protection to both *mothers and fathers*, irrespective of which *parents* [sic] exercises the right of custody or of visitation") (emphasis added).

islators sought to eliminate the incentive of parents who had not prevailed in custody proceedings in one state to abduct their children across state lines to seek relitigation in another forum. Statements of purpose in both the UCCJA and the PKPA indicate a legislative intent to promote the best interests of children involved in custody disputes. Both statutes seek to accomplish this goal by providing clear jurisdictional rules intended to identify the jurisdiction in the best position to decide the merits of the case and by requiring courts to give full faith and credit to custody determinations made by sister states.[14]

Although originally aimed at custody disputes between parents, courts have broadly interpreted the UCCJA and the PKPA to cover many types of custody proceedings, including proceedings to terminate parental rights, guardianship proceedings, and adoption proceedings.[15] However, because these statutes do not contemplate the practical and policy differences between custody disputes in the context of a divorce or separation and custody disputes in the context of an adoption, the application of the UCCJA and the PKPA to interstate adoption proceedings often produces results that are contrary to the stated purposes of these statutes. Part I of this Note examines the practical differences between adoption proceedings and the custody proceedings contemplated by the UCCJA and the PKPA, illustrating how the application of the UCCJA and the PKPA lead to arbitrary jurisdictional results when applied to most interstate challenges to the adoption of infants. Part II examines the policy differences between adoption proceedings and the custody proceedings contemplated by the UCCJA and the PKPA, illustrating how application of the UCCJA and the PKPA does not identify the jurisdiction in the best position to decide interstate adoption challenges. Part III then illustrates how arbitrary jurisdictional determinations in interstate adoption cases can have substantive law effects, despite Congress's characterization of the PKPA as "purely procedural."[16] Part IV offers potential amendments and recommendations designed to make the language and the policies of the UCCJA and the PKPA compatible with interstate adoption cases.

---

[14] *See infra* notes 23–52 and accompanying text.
[15] *See infra* notes 56–72 and accompanying text.
[16] *See infra* note 117.

I. Practical Differences that Distinguish Adoption Custody Cases from Divorce Custody Cases: Interstate Infants Do Not Have "Home States"

A. *The Legal Context*

1. The Problem of the Interstate Child

The Full Faith and Credit Clause of the United States Constitution ensures, in most cases, that judicial orders of one state are enforced in all other states.[17] It does not require, however, that an out-of-state judgment be accorded a degree of finality beyond that given to the judgment in the state where it was rendered. Because child custody orders are in most states modifiable upon proof of changed circumstances,[18] the Supreme Court has ruled that the Full Faith and Credit Clause does not generally foreclose modification of a child custody order of a sister state upon proof of changed circumstances.[19] Prior to the promulgation of uniform legislation, the possibility of modification of custody decrees in another state provided an avenue of recourse for parents unwilling to accept adverse custody judgments. Noncustodial parents could "snatch" their child, move to another state, and try to relitigate the custody issue. Whether due to local chauvinism,[20] different legal standards, or simply different views of the merits, noncustodial parents were often successful in finding a court in a "friendlier forum" willing to switch the custody determination. Of course, if located by the parent originally awarded custody, the child was then susceptible to

---

[17] U.S. Const. art. IV, § 1.

[18] *See* Clark, *supra* note 9, § 19.9 at 836–47.

[19] People of State of New York *ex rel.* Halvey v. Halvey, 330 U.S. 610, 614–15 (1947); Kovacs v. Brewer, 356 U.S. 604, 607–08 (1958). *See also* UCCJA Prefatory Note, 9 U.L.A. 117 (1988):

> [M]any states have felt free to modify custody decrees of sister states almost at random although the theory usually is that there has been a change of circumstances requiring a custody award to a different person . . . . Generally speaking, there has been a tendency to over-emphasize the need for fluidity and modifiability of custody decrees at the expense of the equal (if not greater) need, from the standpoint of the child, for stability of custody decisions once made.

[20] For an illuminating example of the depth of local prejudice in these cases, see Salisbury v. Salisbury, 657 S.W.2d 761, 768 (Tenn. Ct. App. 1983) ("We conclude that the Tennessee decree, as modified by this Court, is entitled to enforcement by the state of Texas, which we would point out would still be a Mexican province had Tennesseans not fought at the Alamo.").

being "snatched back." Thus, child victims of "parental kidnaping" were frequently moved from home to home and state to state in attempts to evade the other parent.[21] Such children were not only deprived of a relationship with the parent originally awarded custody, but they were also denied the stable home environment essential to formative development.[22]

2. The Uniform Child Custody Jurisdiction Act

The search for a legislative solution to interstate custody disputes led to the promulgation of the Uniform Child Custody Jurisdiction Act in 1968 by the National Conference of Commissioners on Uniform State Laws (NCCUSL). The Act was approved by the American Bar Association and recommended to the states for passage that same year.

The UCCJA lists nine statutory purposes,[23] all of which are aimed at remedying the "intolerable state of affairs where self-

---

[21] *See* UCCJA Prefatory Note, 9 U.L.A. 116:

> [Noncustodial parents] will remove the child in an unguarded moment or fail to return him after a visit and will seek their luck in the court of a distant state where they hope to find—and often do find—a more sympathetic ear for their plea of custody. The party deprived of the child may then resort to similar tactics to recover the child and this "game" may continue for years, with the child thrown back and forth from state to state, never coming to rest in one single home and in one community.

[22] *See* GOLDSTEIN, *supra* note 6, at 31–39.

[23] UCCJA § 1 provides:

> (a) The general purposes of this Act are to:
>
> (1) avoid jurisdictional competition and conflict with courts of other states in matters of child custody which have in the past resulted in the shifting of children from state to state with harmful effects on their well-being;
>
> (2) promote cooperation with the courts of other states to the end that a custody decree is rendered in that state which can best decide the case in the interest of the child;
>
> (3) assure that litigation concerning the custody of a child take place ordinarily in the state with which the child and his family have the closest connection and where significant evidence concerning his care, protection, training, and personal relationships is most readily available, and that courts of this state decline the exercise of jurisdiction when the child and his family have a closer connection with another state;
>
> (4) discourage continuing controversies over child custody in the interest of greater stability of home environment and of secure family relationships for the child;
>
> (5) deter abductions and other unilateral removals of children undertaken to obtain custody awards;
>
> (6) avoid re-litigation of custody decisions of other states in this state insofar as feasible;
>
> (7) facilitate the enforcement of custody decrees of other states;
>
> (8) promote and expand the exchange of information and other forms of mutual assistance between the courts of this state and those of other states

help and the rule of 'seize-and-run' prevail rather than the orderly processes of the law."[24] The UCCJA establishes two primary grounds for the valid exercise of subject matter jurisdiction over custody proceedings: the "home state" provision and the "best interest" provision.[25] Neither ground is exclusive nor given absolute priority by the legislative text.[26] "Home state" jurisdiction seeks to identify the state in which a child has spent a substantial amount of time immediately preceding the custody proceedings.[27] "Best interest" jurisdiction, also called "significant

---

concerned with the same child; and

(9) make uniform the law of those states which enact it.

[24] UCCJA Prefatory Note, 9 U.L.A. 117 (1988).

[25] UCCJA § 3, Comment. UCCJA § 3 provides, in relevant part:

(a) A court of this State which is competent to decide child custody matters has jurisdiction to make a child custody determination by initial or modification decree if:

(1) this State (i) is the home state of the child at the time of the commencement of the proceeding, or (ii) had been the child's home state within 6 months before commencement of the proceeding and the child is absent from this State because of his removal or retention by a person claiming his custody or for other reasons, and a parent or person acting as a parent continues to live in this State; or

(2) it is in the best interest of the child that a court of this State assume jurisdiction because (i) the child and his parents, or the child and at least one contestant, have a significant connection with this State, and (ii) there is available in this State substantial evidence concerning the child's present or future care, protection, training, and personal relationships . . . .

Most courts have held that both provisions require that the conditions requisite for jurisdiction exist at the "commencement of the proceedings." *See, e.g.*, Stubbs v. Weathersby, 869 P.2d 893, 896 (Or. Ct. App. 1994); *In re* A.E.H., 468 N.W.2d 190, 200 (Wis. 1991). "Commencement" has been interpreted by most courts to mean the date on which the petition at issue is filed. *See, e.g.*, Umina v. Malbica, 538 N.E.2d 53, 57 (Mass. App. Ct. 1989); Martinez v. Reed, 490 So. 2d 303, 306 (La. Ct. App. 1986).

Two additional grounds for jurisdiction are included in the UCCJA that are not relevant to the current analysis. A state may assert "emergency" jurisdiction if the child is present in the state and has been abandoned or is in danger of abuse. UCCJA § 3(a)(3). A fourth provision allows jurisdiction if no other state meets any of the other three jurisdictional provisions or if all qualifying states have refused to exercise jurisdiction and if it is in the best interest of the child to assume jurisdiction. UCCJA § 3(a)(4).

[26] *See* Houtchens v. Houtchens, 488 A.2d 726, 730 (R.I. 1985) (rejecting argument for application of a "home state" preference based on the policies of the UCCJA). However, some courts have interpreted the UCCJA to express a policy preference for "home state" jurisdiction. *See, e.g.*, Grayson v. Grayson, 454 A.2d 1297, 1299 (Del. Fam. Ct. 1982) ("[W]here litigation is occurring or can occur in the child's 'home state,' such litigation should, in most instances, take precedence over litigation occurring or that might occur in a state whose jurisdiction depends on [the 'best interest' provision].").

[27] "Home state" is defined by UCCJA § 2(5) as

the state in which the child, immediately preceding the time involved, lived with his parents, a parent, or a person acting as a parent, for at least 6 consecutive months, and in the case of a child less than 6 months old a state

connection" jurisdiction, seeks to identify a state that has: (1) a significant connection to the child and at least one party in the custody dispute; and (2) access to substantial evidence concerning the child's present and future care.[28] The physical presence of the child in the state is not a prerequisite to jurisdiction under the UCCJA.[29]

Even if a state satisfies one or both of these jurisdictional provisions, it may not exercise jurisdiction under the Act if a custody case involving the same child is pending in the court of another state operating "substantially in conformity" with the UCCJA.[30] Following resolution of a custody case in another state, courts must recognize and enforce out-of-state decrees promulgated under the Act, unless they have been modified in accordance with the jurisdictional requirements of the UCCJA.[31] A state otherwise meeting the jurisdictional requirements of the Act may properly modify an out-of-state custody determination only if the state that originally rendered the judgment no longer satisfies the jurisdictional requirements of the UCCJA or declines to exercise jurisdiction.[32]

---

in which the child lived from birth with any of the persons mentioned. Periods of temporary absence of any of the named persons are counted as part of the 6-month or other period.

[28] Interpretations vary as to whether the "best interest" language at the outset of this provision establishes a third requirement for the assertion of this jurisdictional ground. *See* CLARK, *supra* note 9, § 12.5, at 466. *Compare In re* Marriage of Hopson, 168 Cal. Rptr. 345, 353 (Cal. App. 3d 1980) ("The first clause of the paragraph is important; jurisdiction exists only if it is in the *child's* interest, not merely the interest or convenience of the feuding parties, to determine custody in a particular state.") (emphasis in original) *with In re* B.B.R., 566 A.2d 1032, 1038 n.18 (D.C. 1989) (analyzing language identical to that in the UCCJA, "The use of the word 'because' at the start of this passage indicates that for the purposes of the PKPA the phrase 'best interest of the child' is defined by the two stated elements.").

[29] UCCJA § 3(c).

[30] UCCJA § 6(a). This "simultaneous proceedings" provision has the effect of establishing exclusive jurisdiction in the first court to properly assume jurisdiction under the Act. *See In re* A.E.H., 468 N.W.2d 190, 204 (Wis. 1991); *In re* B.B.R., 566 A.2d at 1037 n.13. The "substantially" qualification allows for the minor variations of the UCCJA that various states may enact. *See* CLARK, *supra* note 9, § 12.5, at 471. This provision does allow a state to exercise jurisdiction if the court of another state where the matter is pending *chooses* to stay that proceeding. UCCJA § 6(a).

[31] UCCJA § 13.

[32] UCCJA § 14(a) provides:

> If a court of another state has made a custody decree, a court of this State shall not modify that decree unless (1) it appears to the court of this State that the court which rendered the decree does not now have jurisdiction under jurisdictional prerequisites substantially in accordance with this Act or has declined to assume jurisdiction to modify the decree and (2) the court of this State has jurisdiction.

Since its promulgation, the UCCJA has been praised as a powerful instrument in curbing forum shopping by losing parties in custody disputes. However, uniform legislation did not provide a complete solution to the problems of parental child-snatching. Despite the successes of the UCCJA, the push for federal legislative intervention soon followed.[33] Proponents of federal legislation noted that many states had been slow to adopt the UCCJA, creating "safe harbors" for child-snatching parents who sought modifications of custody decrees.[34] Some of the states that had adopted the Act enacted variations of material provisions, undermining uniformity among states.[35] Even among states that had adopted identical language, courts in some states interpreted the same statutory provisions differently from courts in other states.[36] Commentators noted the ample opportunities for forum shopping still available under the UCCJA. Perhaps most notably, the "best interest" jurisdictional ground, based on the ambiguous terms "significant connection" and "substantial evidence,"[37] proved susceptible to abuse.[38] Prior to commencement

---

[33] *See PKA Hearings*, *supra* note 1, at 133 (testimony of Russell M. Coombs, Rutgers University School of Law) ("The Uniform Child Custody Jurisdiction Act is excellent legislation. It is highly desirable that more States enact it . . . . But, no matter how well the States do their job, there is an essential Federal role."); Bob Westgate, *Child-Snatching: The Game Nobody Wins*, SINGLE PARENT, July-Aug. 1979, at 10–11, *reprinted in Parental Kidnaping Hearings*, *supra* note 12, at 123–24.

[34] *See Addendum to PKPA Hearings*, *supra* note 13, at 139 (additional statement of Sen. Wallop) ("This provision . . . should serve . . . as a significant inducement to the 11 states and the District of Columbia which have yet to adopt the Uniform Act, since by doing so their custody and visitation decrees would then be entitled to be recognized, enforced and not modified by sister states."). The UCCJA has now been adopted by all 50 states, the District of Columbia, and the Virgin Islands. *See* UCCJA Table of Jurisdictions, 9 U.L.A. at 115–16.

[35] *See, e.g.*, Custody of Brandon, 551 N.E.2d 506, 508 n.4 (Mass. 1990) (noting that although MASS. GEN. L. ch. 209B was based on the UCCJA, it differs in material respects from the Uniform Act).

[36] *See Parental Kidnaping Hearings*, *supra* note 12, at 100 (statement of Dr. Doris Jonas Freed, Esq., chairperson, A.B.A. Custody Committee); *Reform of the Federal Criminal Laws: Hearings on S. 1722 and S. 1723 Before the Senate Comm. on the Judiciary*, 96th Cong., 1st Sess. 10,670 (1979) [hereinafter *Criminal Reform Hearings*] (statement of Russell M. Coombs, associate professor, Rutgers University School of Law).

[37] UCCJA § 3(a)(2).

[38] *See* CLARK, *supra* note 9, § 12.5, at 467 ("It is not surprising, in view of the imprecision and vagueness of the significant connection provision in the Act, that the cases construing that provision have reached widely diverse results."); *Parental Kidnaping, 1979: Hearing on Examination of the Problem of "Child Snatching" Before the Subcomm. on Child and Human Development of the Senate Comm. on Labor and Human Resources*, 96th Cong., 1st Sess. 51 (1979) [hereinafter *Child-Snatching Hearings*] (testimony of Brigitte M. Bodenheimer, University of California at Davis School of Law) (describing how the "significant connection" provision is being used as a loophole in the UCCJA).

of child custody proceedings, a parent who feared an adverse judgment in the child's "home state" could simply take the child to another state with custody standards more favorable to that parent. Even courts of other states operating under the UCCJA often could be persuaded to assume jurisdiction based on a "significant connection" and "substantial evidence" analysis.[39] In addition, after completion of custody proceedings in one state, the UCCJA modification provision allows another state that meets the Act's jurisdictional requirements to modify the custody decree if the original state no longer has jurisdiction under the UCCJA.[40] Thus, some courts interpreted the UCCJA to allow modification of custody determinations of other states after children had been brought into the state by the noncustodial parent and retained there for six months prior to the filing of the modification proceedings.[41] These courts reasoned that after six months, the state where the original custody determination was rendered was no longer the child's "home state." Modification in such cases thus required only that the original state be found to no longer satisfy the "best interest" provision. Because the point at which a connection is no longer "significant," or at which evidence is no longer "substantial," is largely a matter of judicial discretion, courts could easily find that other states no longer satisfied the jurisdictional requirements of the UCCJA.[42]

---

[39] For a thoughtful critique of the literal interpretation of the UCCJA and PKPA in these circumstances, *see In re* B.B.R., 566 A.2d 1032, 1043 (D.C. 1989) (Schwelb, J., concurring) ("Suppose . . . the Platts had filed a day earlier in the District . . . . Should this court then say that because the child has been brought to the District, and because connections with this jurisdiction have been created for him, the case must be tried in the wrongdoers' lair, three thousand miles from the home of the child's mother? I find such a conclusion so inequitable as to be beyond endurance.").

[40] UCCJA § 14.

[41] *See, e.g., In re* Marriage of Settle, 556 P.2d 962, 968–69 (Or. 1976), *overruled by In re* Custody of Ross, 630 P.2d 353, 363 (Or. 1981). The indeterminacy of the UCCJA as applied to custody disputes between divorcing or separating parents has received extensive treatment by commentators. *See infra* note 55 (citing law review articles on the UCCJA and PKPA).

[42] *See* CLARK, *supra* note 9, § 12.5, at 468 ("Whatever the forces controlling the outcomes of the cases, it seems impossible to advance any reliable generalization concerning the meaning and application of the significant connection test of jurisdiction, other than to say that this test permits a court to accept or reject a finding of jurisdiction on almost any state of facts which is likely to arise.").

Compare In re A.E.H., 468 N.W.2d 190, 201 (Wis. 1991) (interpreting the "best interest" provision to require "optimum access to relevant evidence"); Sullivan v. Sullivan, 451 N.Y.S.2d 851, 853 (N.Y. App. Div. 1982) (noting that the "best interest" provision was not intended to be interpreted so broadly as to confer jurisdiction based on the mere presence in the state of parent and child); *In re* Custody of Bozarth, 538 N.E.2d 785, 792 (Ill. App. Ct. 1989) ("[T]he 'substantial evidence' requirement is not

3. The Parental Kidnaping Prevention Act

Congress enacted the Parental Kidnaping Prevention Act in 1980 to accord full faith and credit to custody decrees.[43] The language of the jurisdictional conditions of the PKPA tracks the language of the jurisdictional bases of the UCCJA, with minor exceptions.[44] The PKPA does not actually confer jurisdiction.

---

intended to be used as a substitute for 'some evidence' but clearly requires a high degree of connection and access to evidence.") *with In re* B.B.R., 566 A.2d at 1039 n.23 ("[A] court is not directed to weigh the relative significance of connections between the child and the competing states. If there is a significant connection to state 1, state 2 must abstain from exercising jurisdiction even if there is a significant connection to it as well."); Grayson v. Grayson, 454 A.2d 1297, 1300 (Del. Fam. Ct. 1982) ("This subsection is not designed to permit a comparison and weighing of factors tying the child to one state or another, a trap that some courts have fallen into."); *In re* Adoption of Child by T.W.C., 636 A.2d 1083, 1089 (N.J. Super. Ct. App. Div. 1994) ("[N]either the UCCJA nor the PKPA requires that one state's 'substantial contacts' be weighed against another state's 'substantial contacts' in order to entrust a single state with jurisdiction."); *In re* Adoption of C.L.W., 467 So. 2d 1106, 1110 (Fla. Dist. Ct. App. 1985) (noting that courts generally recognize existing familial relationships as contacts sufficient to satisfy the "significant connection" provision).

[43] PKPA, § 7(c) provides:

> The general purposes of sections 6 to 10 of this Act are to-
>
> (1) promote cooperation between State courts to the end that a determination of custody and visitation is rendered in the State which can best decide the case in the interest of the child;
>
> (2) promote and expand the exchange of information and other forms of mutual assistance between States which are concerned with the same child;
>
> (3) facilitate the enforcement of custody and visitation decrees of sister States;
>
> (4) discourage continuing interstate controversies over child custody in the interest of greater stability of home environment and of secure family relationships for the child;
>
> (5) avoid jurisdictional competition and conflict between State courts in matters of child custody and visitation which have in the past resulted in the shifting of children from State to State with harmful effects on their well-being; and
>
> (6) deter interstate abductions and other unilateral removals of children undertaken to obtain custody and visitation awards.

See also CLARK, *supra* note 9, § 12.5, at 476; Thompson v. Thompson, 798 F.2d 1547, 1553–54 (9th Cir. 1986), *aff'd*, 484 U.S. 174 (1988) (providing extensive survey of legislative history of the PKPA).

[44] PKPA, § 8(c) establishes the jurisdictional conditions of the PKPA. This section provides, in relevant part:

> A child custody determination made by a court of a State is consistent with the provisions of this section only if-
>
> (1) such court has jurisdiction under the law of such State; and
>
> (2) one of the following conditions is met:
>
> (A) such State (i) is the home State of the child on the date of the commencement of the proceeding, or (ii) had been the child's home State within six months before the date of the commencement of the proceeding and the child is absent from such State because of his removal or retention by a contestant or for other reasons, and a contestant continues to live in such State;

The jurisdictional provisions of the UCCJA are restated in the PKPA as conditions for enforcement of an out-of-state custody decree. However, because a court need not give full faith and credit to out-of-state decrees that do not satisfy these conditions, the PKPA is functionally a jurisdictional statute, and is so labeled in both case law and commentary.[45] Under the PKPA, preference is given to "home state"[46] jurisdiction. Unlike the UCCJA, "best interest" jurisdiction may be asserted in accordance with the PKPA only if no other state qualifies as the child's "home state,"[47] thus closing the jurisdictional loophole created by the alternative bases for jurisdiction in the UCCJA.[48] Like the UCCJA, custody determinations may be modified under the PKPA only if the modifying state has jurisdiction and the original state either no longer has jurisdiction or declines to exercise jurisdic-

---

(B)(i) it appears that no other State would have jurisdiction under subparagraph (A), and (ii) it is in the best interest of the child that a court of such State assume jurisdiction because (I) the child and his parents, or the child and at least one contestant, have a significant connection with such State other than mere physical presence in such State, and (II) there is available in such State substantial evidence concerning the child's present or future care, protection, training, and personal relationships.

The PKPA, like the UCCJA, also contains conditions for "emergency" jurisdiction and for "default" jurisdiction. PKPA §8(c)(1)(C)–(D).

[45] *See, e.g.*, Clark, *supra* note 9, § 12.5, at 479–80; William M. Schur, *Adoption Procedure*, *in* 1 Adoption Law and Practice § 4.02(7), at 4-51 to 4-54 (Joan H. Hollinger ed., 1994); Katherine G. Thompson et al., *Contested Adoptions*, *in* 2 Adoption Law and Practice § 8.06, at 8–70 (Joan H. Hollinger ed., 1994).

In a given case, the PKPA provisions operate in conjunction with the provisions of the applicable state version of the UCCJA. *See* Clark, *supra* note 9, § 12.5, at 464 ("[A] logical analysis of the statutes and case law is extremely difficult and in addition is likely to require the lawyer or court to move, step by step, from one of the Acts to the other, and from one of the provisions in one of the Acts to other provisions in that Act."). However, due to Supremacy Clause considerations, the jurisdictional requirements of the PKPA preempt any conflicting provisions of the UCCJA. *See, e.g.*, *In re* Adoption of Child by T.W.C., 636 A.2d at 1088; *In re* B.B.R., 566 A.2d at 1036 n.10; Adoption of Zachariah K., 8 Cal. Rptr. 2d 423, 429 (Cal. Ct. App. 1992).

[46] "Home state" is defined by PKPA § 8(b)(4) as:

[H]ome State means the State in which, immediately preceding the time involved, the child lived with his parents, a parent, or a person acting as a parent, for at least six consecutive months, and in the case of a child less than six months old, the State in which the child lived from birth with any of such persons. Periods of temporary absence of any of such persons are counted as part of the six-month or other period.

[47] *See, e.g.*, *In re* B.B.R., 566 A.2d at 1037 ("Thus, if any state qualifies as the 'home State,' that state has exclusive jurisdiction over custody matters."); *In re* Adoption of Child by T.W.C., 636 A.2d at 1087; E.E.B. v. D.A., 446 A.2d 871, 879 (N.J. 1982).

[48] *See* CLARK, *supra* note 9, § 12.5, at 480; *see also supra* notes 37–43 and accompanying text.

tion under the Act.[49] However, in addition to the grounds for jurisdiction enumerated in the UCCJA, the PKPA contains a provision for "continuing jurisdiction."[50] According to this provision, once a custody determination is rendered in conformity with the PKPA, the state that renders that determination retains jurisdiction as long as the state has jurisdiction under its own law and the child or any party to the custody dispute remains a resident of the state.[51] To the extent that this provision expands the scope of circumstances under which a court has jurisdiction beyond the scope of jurisdiction under the UCCJA, modification will prove more difficult under the PKPA scheme.[52]

The UCCJA and the PKPA have not eliminated all jurisdictional indeterminacy from interstate child custody cases. Courts continue to interpret the PKPA and the UCCJA creatively to escape the restraining effects of the jurisdictional directives of the statutes.[53] Thus, forum court favoritism continues to play a

---

[49] PKPA § 8(f).

[50] PKPA § 8(c)(2)(E).

[51] PKPA § 8(d) provides:

> The jurisdiction of a court of a State which has made a child custody determination consistently with the provisions of this section continues as long as the requirement of subsection (c)(1) of this section continues to be met and such State remains the residence of the child or of any contestant.

[52] The phrase "under the law of such State" in subsection (c)(1) of the PKPA has received alternative interpretations. If interpreted as an explicit reference to that state's version of the UCCJA, the "continuing jurisdiction" provision does not make modification of an out-of-state custody decree any more difficult. No state could be held to have "continuing jurisdiction" that did not have jurisdiction under the UCCJA jurisdictional provision, and thus the PKPA "continuing jurisdiction" provision would add nothing to the statutory scheme. *See* Clark, *supra* note 9, § 12.5, at 483; *In re* A.E.H., 468 N.W.2d 190, 209 (Wis. 1991) (holding that since California no longer meets the jurisdictional requirements of the UCCJA, it no longer has jurisdiction under its own law, and modification under the PKPA is appropriate); Serna v. Salazar, 651 P.2d 1292, 1295 (N.M. 1982) (rejecting the interpretation that a court has continuing jurisdiction regardless of UCCJA jurisdiction as long as one of the parties remains a resident of the state). If, however, "under the law of such State" is interpreted to mean jurisdiction under any state law, such as the state's divorce statutes or adoption statutes, modification under the PKPA becomes substantially more difficult. *See In re* Adoption of Zachariah K., 8 Cal. Rptr. 2d 423, 429 (Cal. Ct. App. 1992) (noting that a major difference between the UCCJA and the PKPA is greater flexibility to modify custody awards under the UCCJA); *In re* Clausen, 502 N.W.2d 649, 658 n.25 (Mich. 1993) (noting the trend of PKPA cases to interpret continuing jurisdiction expansively).

[53] *See, e.g.*, E.E.B. v. D.A., 446 A.2d 871, 877 (N.J. 1982) (holding that the failure of an Ohio court to conduct a best interests hearing before ordering the return of a child to biological parents constituted a refusal to exercise jurisdiction to modify the custody decree, thus allowing the New Jersey court to modify that custody decree under the PKPA).

role in many of these disputes.[54] Nevertheless, the statutory scheme created by the UCCJA and the PKPA does provide a uniform mechanism for resolution of jurisdictional disputes in child custody cases. When applied to custody disputes between parents, such as those following divorce or separation, application of the UCCJA and the PKPA should result in adjudication of the case in the forum in the best position to decide the merits and will normally deny a parent any tactical advantage formerly gained by "snatching" the child from a custodial parent.[55]

B. *Application to Interstate Adoption Proceedings*

1. The "Custody Determination" Definition

Neither the UCCJA nor the PKPA contains any reference to adoption proceedings.[56] In fact, the legislative history of the PKPA does not discuss the application of the statute to adoptions or prospective adoptive parents.[57] However, numerous courts have held both the UCCJA and the PKPA applicable to jurisdictional issues in guardianship proceedings,[58] proceedings to terminate

---

[54] *See* CLARK, *supra* note 9, § 12.5, at 494. ("It is impossible to tell whether the enactment of the UCCJA and the PKPA have [sic] brought about a diminution of the courts' local chauvinism which has been such a troublesome aspect of custody litigation.").

[55] There has been much written questioning the success of the UCCJA and the PKPA with regard to the application of the statutes to custody disputes between parents. *See, e.g.*, CLARK, *supra* note 9, § 12.5, at 494; *see also* Anne B. Goldstein, *The Tragedy of the Interstate Child: A Critical Reexamination of the Uniform Child Custody Jurisdiction Act and Parental Kidnaping Prevention Act*, 25 U.C. DAVIS L. REV. 845 (1992); James C. Murray, *One Child's Odyssey Through the Uniform Child Custody Jurisdiction and Parental Kidnaping Prevention Acts*, 1993 WIS. L. REV. 589.

[56] However, some of the state versions of the UCCJA do explicitly either include or exclude adoptions from the statutory definition of "custody proceeding." *See* D.C. CODE ANN. § 16-4502(2) (1993) (including adoptions); GA. CODE ANN. § 19-9-42(3) (1994) (including adoptions); MONT. CODE ANN. § 40-7-103(3) (1993) (including adoptions); N.H. REV. STAT. ANN. § 458-A:2(III) (1993) (explicitly excluding adoptions, proceedings to terminate parental rights, and guardianship proceedings); N.Y. DOM. REL. LAW § 75-c(3) (McKinney 1995) (explicitly excluding adoptions, proceedings to terminate parental rights, and guardianship proceedings).

[57] The only reference to adoption proceedings in the hearings on the PKPA is found in an explanation of the major provisions of the Draft of the Hague Convention, a multilateral child abduction treaty similar to the UCCJA. *Addendum to PKPA Hearings*, *supra* note 13, at 348 (stating that the Convention applies to "a prospective adoptive parent . . . who removes or retains a child in order to assume the child's care and control"). This document is included, without further explanation, under the "Contents" heading of "Letters, Articles, and Miscellaneous Items." *Id.* at III.

[58] *See, e.g.*, *In re* A.E.H., 468 N.W.2d 190, 194, 199 (Wis. 1991); Gribkoff v. Bedford, 711 P.2d 176, 177–78 (Or. Ct. App. 1985); *In re* Guardianship of Walling, 727 P.2d 586, 589 (Okla. 1986).

parental rights,[59] and adoption proceedings.[60] The statutory definition of "custody determination" in the UCCJA[61] is somewhat different from that in the PKPA.[62] However, both definitions apply to court orders "providing for" the custody of a child. As several courts have noted, proceedings that terminate parental rights and proceedings that grant adoption decrees literally fit within the scope of these definitions because they ultimately "provide for" the custody of the child at issue; some courts have even referred to adoption as the "ultimate" custody proceeding.[63] The Official Comment to the UCCJA section on definitions indicates that "custody proceeding" should be interpreted broadly.[64]

---

[59] *See, e.g., In re* A.E.H., 468 N.W.2d at 194, 200; White v. Blake, 859 S.W.2d 551, 561 (Tex. Ct. App. 1993); *In re* Steven C., 486 N.W.2d 572, 573–74 (Wis. Ct. App. 1992); *In re* L.C., 857 P.2d 1375, 1377 (Kan. Ct. App. 1993). *But see In re* Johnson, 415 N.E.2d 108, 110 (Ind. Ct. App. 1981) (holding termination of parental rights not to be encompassed by the UCCJA, but properly determined in accordance with the adoption statutes).

[60] *See, e.g.*, State *ex rel.* Torres v. Mason, 848 P.2d 592, 595 (Or. 1993); *In re* Adoption of Baby Girl B., 867 P.2d 1074, 1077–78 (Kan. Ct. App. 1994); *In re* Adoption of Child by T.W.C., 636 A.2d 1083, 1086–87 (N.J. Super. Ct. App. Div. 1994); E.E.B. v. D.A., 446 A.2d 871, 878 (N.J. 1982); Gainey v. Olivo, 373 S.E.2d 4, 6 (Ga. 1988); *In re* Steven C., 486 N.W.2d at 573–74; *In re* Adoption of Zachariah K., 8 Cal. Rptr. 2d 423, 428 (Cal. Ct. App. 1992); *In re* Clausen, 502 N.W.2d 649, 656 (Mich. 1993). *See also* Foster v. Stein, 454 N.W.2d 244, 246–47 (Mich. Ct. App. 1990) (holding "custody proceeding" to include adoptions despite legislation that had recently deleted explicit reference to adoptions from the definition of "custody proceeding" in the Michigan UCCJA).

[61] UCCJA § 2 provides, in relevant part:

> (2) "custody determination" means a court decision and court orders and instructions providing for the custody of a child, including visitation rights; it does not include a decision relating to child support or any other monetary obligation of any person;
>
> (3) "custody proceeding" includes proceedings in which a custody determination is one of several issues, such as an action for divorce or separation, and includes child neglect and dependency proceedings;
>
> (4) "decree" or "custody decree" means a custody determination contained in a judicial decree or order made in a custody proceeding, and includes an initial decree and a modification decree; . . . .

[62] PKPA § 8(b)(3) provides:

> "custody determination" means a judgment, decree, or other order of a court providing for the custody or visitation of a child, and includes permanent and temporary orders, and initial orders and modifications.

[63] *See* State *ex rel.* Torres, 848 P.2d at 593; *In re* Johnson, 415 N.E.2d at 110; *In re* Steven C., 486 N.W.2d at 573; *In re* Adoption of Baby Girl B., 867 P.2d at 1078.

[64] UCCJA § 2 Commissioners' Comment.

However, some proceedings that technically fit within the scope of the UCCJA have been found to involve unique policy considerations that merit a less than strict application of the statute. *See In re* Welfare of Mullins, 298 N.W.2d 56, 59–60 (Minn. 1980) (noting the importance of the analytical distinction between custody decrees arising from divorce and custody decrees arising from dependency); *see generally* Danny R. Veilleux, Annotation, *What Types of Proceedings or Determinations Are*

The UCCJA definition itself explicitly includes neglect and dependency proceedings,[65] the most common grounds for the involuntary termination of parental rights.[66] In writings subsequent to the promulgation of the Model Act, the late Professor Brigitte M. Bodenheimer, the Reporter for the UCCJA, stated that adoption proceedings and proceedings to terminate parental rights should be covered by the UCCJA.[67]

In addition to arguments from the statutory language, courts applying the UCCJA and the PKPA to adoption proceedings have found support in the statute's stated purposes. Some courts have found application of the jurisdictional provisions of the UCCJA and the PKPA to adoption proceedings to be "consistent" with reduced competition among states, increased cooperation among states, and better identification of the jurisdiction with the closest connection to the child.[68]

Most commentators agree that the UCCJA and the PKPA should apply to adoption proceedings.[69] The line of argument advanced by Professor Homer H. Clark, Jr., on this issue in his family law treatise is especially noteworthy. After acknowledging the state of confusion of the jurisdictional law, and after analyzing the issues and interests relevant in adoption cases, Professor Clark concludes that subject matter jurisdiction in adoption cases should exist "where a) the prospective adoptive parents, the petitioners, reside in the jurisdiction, and b) the child is physically present in the jurisdiction."[70] He then observes that his analysis is "in the process of being adopted" through the "vehicles" of the UCCJA and the PKPA.[71] As the following analysis will illustrate,

---

Governed by the Uniform Child Custody Jurisdiction Act (UCCJA) or the Parental Kidnaping Prevention Act (PKPA), 78 A.L.R. 4th 1028 (1990).

[65] UCCJA § 2(3).

[66] *See infra* text accompanying note 105.

[67] Brigitte M. Bodenheimer & Janet Neeley-Kvarme, *Jurisdiction Over Child Custody and Adoption After Shaffer and Kulko*, 12 U.C. DAVIS L. REV. 229, 252–53 (1979). *See infra* note 139 and accompanying text.

[68] *See* State *ex rel.* Torres, 848 P.2d 592, 594 (Or. 1993); White v. Blake, 859 S.W.2d 551, 562–63 (Tex. Ct. App. 1993); Gainey v. Olivo, 373 S.E.2d 4, 6 (Ga. 1988).

[69] *See, e.g.*, CLARK, *supra* note 9, § 20.3, at 873–74; Schur, *supra* note 45, § 4.02(6)–(7) at 4-48 to 4-49, 4-51, 4-55; Thompson, *supra* note 45, § 8.06, at 8–69. Such conclusions seem grounded much more in the need for *some* solution to jurisdictional problems in adoption cases than in whether the UCCJA and the PKPA provide the *correct* solution to these problems. In this regard, the title of the only previous law review article directly to address the subject is particularly illuminating. *See* Bernadette W. Hartfield, *The Uniform Child Custody Jurisdiction Act and the Problem of Jurisdiction in Interstate Adoption: An Easy Fix*, 43 OKLA. L. REV. 621 (1990).

[70] CLARK, *supra* note 9, § 20.3, at 872–73.

[71] *Id.* § 20.3, at 873.

Professor Clark would find upon closer inspection that the actual operation of the jurisdictional provisions of the UCCJA and the PKPA in adoption cases is seldom consistent with his analysis.[72]

2. The "Home State" Definition

There are far more couples in the United States waiting to adopt an infant than there are infants available for adoption. According to one estimate, over one million couples vie for the 30,000 white infants available for adoption each year.[73] Adoption

---

[72] *See infra* note 147. There is good reason for confusion about the effects of the application of the UCCJA and the PKPA to adoption cases. *See* Schur, *supra* note 45, at 134 (Supp. 1994) ("[C]ourts remain puzzled about precisely how the UCCJA and PKPA determine which state may exercise jurisdiction either to issue an adoption order or modify a custody order issued by another state.").

[73] *See* Cynthia Crossen, *Hard Choices: In Today's Adoptions, the Biological Parents Are Calling the Shots*, WALL ST. J., Sept. 14, 1989, at 1 (citing estimate by the National Committee for Adoption).

The federal government stopped collecting information on adoptions in 1975. *See* Michael Rezendes, *Should Race Matter in Adoptions?*, NORTHERN N.J. REC., Mar. 27, 1994, at L1. Thus, reliable adoption statistics are not readily available. *See* Mary Richter-Zennik, *How to Dispel Adoption Myths*, BLOOMINGTON PANTAGRAPH, Dec. 13, 1994, at C1. One statistic often cited is that for every available child in the United States there are 40 couples who want to adopt. *See, e.g.*, Maudlyne Ihejirika, *From Belarus With Love: Families Find an Adoption Connection*, CHI. SUN-TIMES, Dec. 14, 1994, at 7. This figure was first published in conjunction with a 1985 study by the National Committee for Adoption. *See* Spencer Rich, *Demand for Adoptions Outnumbers the Babies: Private Study Is First in Decade*, WASH. POST, Nov. 20, 1985, at A19. However, this statistic may be misleading, since it used the number of couples who had visited fertility clinics in 1982 (2 million) as the source of its number of couples "waiting to adopt." *See* Marlys Harris, *Where Have all the Babies Gone? (Adoption of Children)*, MONEY, Dec. 1, 1988, at 164. A 1990 study by the National Center for Health Statistics found that only 200,000 women tried to adopt a child in 1988. *See* Philip J. Hilts, *New Study Challenges Estimates on Odds of Adopting a Child*, N.Y. TIMES, Dec. 10, 1990, at B1; *see generally* CHRISTINE A. BACHRACH ET AL., U.S. DEP'T OF HEALTH AND HUMAN SERVS., ADVANCE DATA NO. 181, ADOPTION IN THE 1980'S (1990). Estimates of the total number of adoptions in the United States each year range from 100,000 to 142,000. However, about half of these adoptions are adoptions by stepparents or relatives. *See* CLARK, *supra* note 9, § 20.1, at 852–53. All estimates agree that the number of healthy infants available for adoption in the United States is very low, probably between 20,000 and 30,000. *See* Bob Sector, *Longstanding Taboos Tossed Aside: Couples Seeking to Adopt a Baby Use Classified Ads*, L.A. TIMES, May 24, 1987, at 1; Susan Everly-Douze, *Parents Find Children Half a World Away*, TULSA WORLD, May 8, 1994, at N1; Kevin Sullivan, *Wanted: Baby for Loving Home; Adoption Through Classifieds Gains Popularity and Critics*, WASH. POST, Oct. 5, 1992, at A1. Thus, even using the most conservative estimates, the ratio of hopeful couples to available infants approaches 7 to 1. It is not uncommon for a white family to spend five years on the waiting list of an adoption agency. *See* Martha Shirk, *Racial Issues Pushed Into Adoption Limelight: Placement of Black Babies Lags Behind That of Whites*, ST. LOUIS POST DISPATCH, July 17, 1994, at 1A; CLARK, *supra* note 9, § 20.1, at 852 ("[I]t is clear that childless couples wishing to adopt a newborn child face long waiting lists or outright rejection in places where the adoption agencies have so many applications that they no longer accept new ones.").

It should be noted that the references above to "white" children and families reflect

procedures in the United States are almost entirely governed by state law.[74] However, since the supply of children available for adoption within a state rarely meets the demand,[75] couples seeking to adopt newborn infants often engage in a national search. The Interstate Compact on the Placement of Children (ICPC) has sought to fill the need for a procedural framework to protect the interests of children involved in these interstate adoptive arrangements.[76]

The vast majority of families that seek to adopt non-relatives in the United States seek to adopt infants.[77] Infants available for

---

the plight of minority children, who, along with older children and children with mental or physical disabilities, are often classified as "hard-to-place." According to statistics from the Child Welfare League of America, African Americans make up only 12% of the population, but African American children make up 42% of the pool of children currently awaiting permanent placements. Transracial adoption initiatives have surfaced as a result of the shortage of available minority adoptive families. As a result, racial matching policies of adoption agencies have been a source of much recent debate. *See* Maria E. Salmine, *For the Sake of the Child: Moving Toward Uniformity in Adoption Law*, 69 WASH. L. REV. 841, 859 (1994); Kathy S. Stale, *Statistics on Adoption in the United States*, FUTURE OF CHILDREN, Spring 1993, at 26, 29.

[74] *See* Dr. Mitchell Wendell & Betsey R. Rosenbaum, *Interstate Adoptions: The Interstate Compact on the Placement of Children*, *in* 1 ADOPTION LAW AND PRACTICE, Appendix § 3-A.01, at 3A-1 (Joan H. Hollinger ed., 1994).

[75] *See* H. JOSEPH GITLIN, ADOPTIONS: AN ATTORNEY'S GUIDE TO HELPING ADOPTIVE PARENTS § 12:02 (1987).

[76] The Council of State Governments proposed the ICPC in 1960. By January of 1990, all 50 states, the District of Columbia, and the Virgin Islands had become signatories. In order to provide the same services and protections for interstate placements that children would have been provided had their placements been intrastate, the Compact prohibits interstate preadoptive placement until the appropriate agency in the state where the child will be placed certifies approval of the arrangement. *See generally* GITLIN, *supra* note 75, § 12:01; Wendell & Rosenbaum, *supra* note 74, § 3-A; Julius Libow, *The Interstate Compact on the Placement of Children—A Critical Analysis*, JUV. & FAM. CT. J., Summer 1992, at 19.

The ICPC has caused additional confusion with regard to interstate jurisdictional questions. One provision of the Compact directs that "the sending agency shall retain jurisdiction over the child . . . until the child is adopted." *See, e.g.*, MINN. STAT. § 257.40 art. 5(a) (1995). Since a court is only one of many parties that may be a "sending agency" under the ICPC, "jurisdiction" arguably has a different meaning under the ICPC than under the UCCJA and PKPA. One noted commentator has argued that under the ICPC, jurisdiction means "responsibility for a child's well-being," while under the UCCJA and PKPA, jurisdiction means "authority to adjudicate the case." Joan H. Hollinger, *Interstate Adoptions: The Interstate Compact on the Placement of Children*, *in* 1 ADOPTION LAW AND PRACTICE 100 (Joan H. Hollinger ed., Supp. 1994) (citing *In re* Zachariah Nathaniel K., 8 Cal. Rptr. 2d 423, 431 (Cal. Ct. App. 1992)). However, judicial interpretations of the ICPC have not always drawn that distinction; many courts have attached jurisdictional importance (in the "adjudicatory authority" sense) to ICPC provisions. *See In re* Baby Girl *(blank)*, 850 S.W.2d 64, 69 (Mo. 1993) (en banc); *In re* Adoption of K.S., 581 A.2d 659, 663 (Pa. Super. Ct. 1990); *In re* Welfare of Mullins, 298 N.W.2d 56, 60 (Minn. 1980); Department of Social Services of Denver v. District Court of the Eighteenth Judicial Dist., 742 P.2d 339, 341–42 (Colo. 1987).

[77] *See supra* note 73 (describing differential in demand between healthy infants and special-needs children).

adoption are placed with adoptive parents as soon as possible after birth. An influential work by a group of experts in child psychology and the law introduced the concept of the "psychological parent-child relationship" as an integral part of a childhood environment of stability and continuity.[78] Early placement provides the best opportunity for the development of such a relationship between a child and adoptive parents.[79] Consistent with this interest, adoption procedures in most jurisdictions allow temporary grants of custody to prospective adoptive parents well before the entrance of the adoption decree.[80]

Early adoptive placement poses a major obstacle to the application of the UCCJA and the PKPA. The definitions of "home state" in statutes are practically identical. When these definitions are applied to infant adoption cases, few of the children involved will be old enough to have lived in a state "immediately preceding the time involved . . . for at least six consecutive months."[81] Thus, in order for a child less than six months old at the commencement of the proceedings to satisfy the definition of "home state" under the UCCJA and the PKPA, the child must have "lived from birth" in a state with a parent or a person acting as a parent.[82] Almost by definition, an infant involved in interstate adoption proceedings will not satisfy that condition: a child that is born to biological parents in one state and then is shortly thereafter placed with prospective adoptive parents in another state has not "lived from birth" in either state. Most courts to address the issue have interpreted the "home state" definition in this literal manner.[83] Most adopted infants thus have no "home

---

[78] GOLDSTEIN, *supra* note 6, at 17–20.

[79] *See id.* at 22.

[80] *Id.* at 35–37, 45–46. Goldstein, Freud, and Solnit favor not only a policy of early placements in adoption (even before birth), but also recommend that adoption decrees be made final upon placement with the adoptive family. They label "waiting periods" as periods of uncertainty that hamper the likelihood of the formation of a meaningful psychological parent-child relationship between the child and the adoptive parents.

[81] *See supra* notes 27, 46.

[82] A few states have enacted a variation of the "home state" definition in their versions of the UCCJA that allows "home state" jurisdiction in the state where a child less than six months old has "resided with any of such persons for a majority of the time since birth." *See* N.H. REV. STAT. ANN. § 458-A:2(V) (1993); N.Y. DOM. REL. LAW § 75-c(5) (McKinney 1995); V.I. CODE ANN. tit. 16, § 116(e) (1994).

[83] *See In re* Adoption of Child by T.W.C., 636 A.2d 1083, 1088 (N.J. Super. Ct. App. Div. 1994); *In re* Adoption of Baby Girl B., 867 P.2d 1074, 1079–80 (Kan. Ct. App. 1994); *In re* Adoption of Zachariah K., 8 Cal. Rptr. 2d 423, 430 (Cal. Ct. App. 1992); *In re* B.B.R., 566 A.2d 1032, 1038 (D.C. 1989); Rogers v. Platt, 245 Cal. Rptr. 532, 538 (Cal. Ct. App. 1988); Rogers v. Platt, 641 F. Supp. 381, 387 (D.D.C. 1986); *In re* Cifarelli, 611 A.2d 394, 397 (Vt. 1992); *See also In re* Clausen, 502 N.W.2d 649,

state" under the UCCJA and the PKPA. As a result, no state can exercise jurisdiction over such a case under the UCCJA or seek enforcement of an order under the PKPA based on the "home state" jurisdictional provision.

The lack of a "home state" does not end the jurisdictional inquiry under the UCCJA or the issue of full faith and credit under the PKPA. The PKPA expressly provides for "best interest" jurisdiction in cases in which no state satisfies the "home state" provision.[84] Under the UCCJA, "best interest" jurisdiction serves as an alternative to "home state" jurisdiction regardless of whether the child has a "home state."[85] Thus, at least one state will be able to assert jurisdiction consistent with the provisions of the UCCJA and the PKPA in all adoption cases.[86] However, as noted above, one of the major accomplishments of the PKPA, as stated in the legislative history, was the restriction of the applicability of the "best interest" provision, which had developed into a jurisdictional loophole under the UCCJA. To the extent that the "home state" provision proves inapplicable to adoption cases, the PKPA policy to limit the opportunity for abuse of the "best interest" provision is thwarted. In nearly all interstate adoption cases, both the state of the natural parents and the state of the adoptive parents could be found to satisfy the "significant connection" and the "substantial evidence" re-

---

673–74 (Mich. 1993) (Levin, J., dissenting). Arguments for recognition of a "functional" home state in cases where infants have lived *almost* from birth in a state have generally been rejected. *See, e.g., In re* B.B.R., 566 A.2d at 1038 n.16 ("The purpose of the statute . . . is to establish clear guidelines for determination of jurisdiction and it is not illogical to think that the drafters had in mind only a situation where the child was born and continued to live in the state of birth up until the relevant events occurred."); *Rogers*, 641 F. Supp. at 385 ("The intent of the PKPA is to leave no room for courts to balance the percentage of time spent or type of living quarters used in one state versus another state, in deciding which state has home state jurisdiction."); *see also* CLARK, *supra* note 9, § 12.5, at 465 ("This section of the statute is sufficiently precise to be applied with little controversy."). *But see In re* Adoption of C.L.W., 467 So. 2d 1106, 1110 (Fla. Dist. Ct. App. 1985) (concluding that a child can live "from birth" in a state in which the child was not present during the first three days of life); *In re* Clausen, 502 N.W.2d at 658 (concluding—without discussion—that "Iowa was unquestionably the home state of the child," where the infant lived for only two weeks in Iowa with state-appointed caregivers before placement in Michigan); *In re* Adoption of K.S., 581 A.2d 659, 662 (Pa. Super. Ct. 1990) (concluding—without discussion—that Delaware is "most certainly the 'home state'" under the UCCJA, in a case involving the adoption of an infant born to a Delaware resident in a Pennsylvania hospital).

[84] *See* PKPA § 8(c)(2)(B).

[85] *See* UCCJA § 3(a).

[86] If the "best interest" provision is not satisfied, PKPA § 8(c)(2)(D) and UCCJA § 3(a)(4) provide a basis for jurisdiction when no state has adequate contacts to meet any of the other provisions.

quirements, especially given the imprecision of these terms.[87] Though either state could be found to have jurisdiction under the UCCJA and the PKPA, only one state may exercise jurisdiction at a time. Provisions in both the UCCJA and the PKPA prevent the exercise of jurisdiction by a state while proceedings concerning the same child are pending in another state.[88] Thus, because the "home state" concept, as drafted, does not accommodate the practical elements of interstate infant adoptions, infant adoption cases under the UCCJA and the PKPA are functionally governed by a "race-to-the-courthouse" rule of jurisdiction. In short, when the UCCJA or the PKPA is applied to the typical interstate infant adoption case, neither state will satisfy "home state" jurisdiction and both states will satisfy "best interest" jurisdiction. Since only one state may properly exercise jurisdiction, the first to file will be the winner of the jurisdictional battle.[89]

A first-to-file jurisdictional rule is arbitrary. The application of the UCCJA and the PKPA to infant adoption cases is thus to a significant degree inconsistent with the policy of both statutes to provide clear jurisdictional directives that discourage jurisdictional competition and conflict. However, these statutes do serve other purposes. In a custody dispute between parties that are asserting identical legal interests, an arbitrary jurisdictional preference for the state whose proceedings were first in time may be perfectly consistent with policy interests in identifying the most appropriate forum, much like a rule that "breaks a tie" by coin toss. However, unlike most custody disputes between parents, jurisdictional disputes in interstate adoption cases do not

---

[87] *See supra* notes 37–42 and accompanying text.

[88] UCCJA § 6(a); PKPA § 8(g).

[89] *Cf. In re* B.B.R., 566 A.2d 1032, 1041 (D.C. 1989) (Schwelb, J., concurring) ("My colleagues hold that the California courts have jurisdiction, and the District of Columbia courts do not, because Ms. Rogers won the race to the courthouse door one day. I agree that the PKPA so provides, but I believe we should decide the issue on broader and less fortuitous grounds than that.").

Where two states could potentially properly exercise jurisdiction, this "race-to-the-courthouse" effect was recognized at the time of the drafting of the PKPA as a necessary evil. *See* Russell M. Coombs, *The "Snatched Child" Is Halfway Home in Congress*, 11 FAM. L.Q. 407, 421 (1978) ("While such a provision might encourage a race to commence proceedings in different states where two states have jurisdiction, the absence of such a provision might encourage not only that kind of race but also a race between litigants and states to press the proceedings to the making of at least temporary orders."). However, the establishment of exclusive jurisdiction for any state that satisfies the "home state" provision confined the number of divorce custody cases in which alternative jurisdiction might exist to the exceptional cases. As illustrated, in infant adoption cases, alternative jurisdiction is the rule.

involve identical legal interests. In adoption cases, the nature of the interests and connections asserted by biological parents are distinct from the interests and connections asserted by prospective adoptive parents. In such cases, a first-to-file rule is not only arbitrary, it is also irrational.

II. Policy Differences that Distinguish Adoption Custody Cases from Divorce Custody Cases: The Interests of the Child Are Largely Irrelevant in Adoption Custody Cases

Both the UCCJA and the PKPA are ultimately motivated by a desire to promote the best interests of children involved in interstate custody disputes.[90] Most child custody disputes, such as custody proceedings following a divorce or separation, involve a custody award between two parents. In such cases, the best interests policy of the UCCJA and the PKPA is consistent with the determinative legal standard applied in the majority of states.[91] Child custody disputes involving adoptions, on the other hand, involve considerations other than the best interests of the child, because adoptions require the termination of parental rights.[92] Thus, in addition to problems resulting from ill-suited textual language, the application of the jurisdictional directives of the UCCJA and the PKPA to interstate adoption proceedings proves arbitrary because these statutes do not contemplate the unique policy considerations relevant to cases involving the termination of parental rights.

A. *The Legal Context*

1. Nature of Parental Rights

The Supreme Court has recognized the fundamental liberty interests of natural parents in the care and custody of their

---

[90] *See supra* notes 23, 43; *see also* UCCJA Prefatory Note, 9 U.L.A. at 118 ("If [interstate cooperation] can be achieved, it will be less important *which* court exercises jurisdiction but that courts of the several states involved act in partnership to bring about the best possible solution for a child's future."); PKPA § 7(a)(4) (noting that "harm to the welfare of children" results from lack of interstate cooperation).

[91] *See* Clark, *supra* note 9, § 19.4, at 797.

[92] *See infra* notes 102–107 and accompanying text.

children.[93] A mere biological connection, without more, does not suffice to trigger parental rights;[94] however, a biological connection does provide the potential for the development of a parent-child relationship.[95] The Supreme Court has held the involuntary termination of parental rights violative of the Due Process Clause of the Fourteenth Amendment absent clear and convincing evidence of parental unfitness.[96] Absent proof of parental unfitness, a child's interests are presumed to be in harmony with the interests of the natural parents.[97] A child has no right to be adopted; thus, a state cannot take a child from natural parents just because the child would be "better off" with another family.[98]

---

[93] *See, e.g.*, Lassiter v. Department of Social Services, 452 U.S. 18, 27 (1981); Quilloin v. Walcott, 434 U.S. 246, 247–48 (1978); Smith v. Organization of Foster Families for Equality and Reform, 431 U.S. 816, 842 (1977); Stanley v. Illinois, 405 U.S. 645, 651 (1972); *see also* Santosky v. Kramer, 455 U.S. 745, 768 (1982).

[94] *See* Lehr v. Robertson, 463 U.S. 248, 262 (1983); *see also Smith*, 431 U.S. at 844 ("[T]he importance of the familial relationship . . . stems from the emotional attachments that derive from the intimacy of daily association, and from the role it plays in 'promot[ing] a way of life' through the instruction of children, as well as from the fact of blood relationship." (quoting Wisconsin v. Yoder, 406 U.S. 205, 231–33 (1972))); Caban v. Mohammed, 441 U.S. 380, 397 (1979) (Stewart, J., dissenting) ("Parental rights do not spring full-blown from the biological connection between parent and child. They require relationships more enduring.").

[95] *See* Lehr v. Robertson, 463 U.S. 248, 262 (1983) ("If [the biological parent] . . . accepts some measure of responsibility for the child's future, he may enjoy the blessings of the parent-child relationship and make uniquely valuable contributions to the child's development."); David L. Chambers, *Rethinking the Substantive Rules for Custody Disputes in Divorce*, 83 MICH. L. REV. 477, 499 (1984) ("Legislators may believe in a mysterious bond between children and their biological parents that will serve the child's interests. At bottom, however, the presumption probably rests on an unexpressed but wholly defensible desire to protect the interests of adult citizens in bearing and raising their own children.").

Society still places a special value on the biological link to offspring, as the rise of new reproductive technologies indicates. *See, e.g.*, Joan H. Hollinger, *From Coitus to Commerce: Legal and Social Consequences of Noncoital Reproduction*, 18 U. MICH. J.L. REF. 865, 874 (1985) ("Much of this demand [for noncoital reproduction] follows from the social and psychological importance people attach to the ideal of having children who are genetically theirs.").

[96] *See Santosky*, 455 U.S. at 767–69; *Quilloin*, 434 U.S. at 255; *see also* Schur, *supra* note 45, § 4.07(1), at 4-122 (noting that "federally protected rights are implicated in every adoption and preadoption suit in which parental rights are terminated"); *In re* Adoption of Baby Boy W., 701 S.W.2d 534, 544–45 (Mo. Ct. App. 1985) (stating that natural father is entitled to presumption of fitness rebuttable only by proof of abandonment or neglect).

[97] *See, e.g.*, *In re* Stevens, 652 A.2d 18, 26 (Del. 1995) ("While it is true that Kelly has certain fundamentally-vested interests, also deserving of constitutional protection, those interests are circumscribed by the competing constitutional rights of her biological parents."); *In re* Clausen, 502 N.W.2d 649, 652, 665–66 (Mich. 1993); *In re* Petition of Otakar Kirchner, 649 N.E.2d 324, 338–39 (Ill. 1995); *see also* Faupel, *supra* note 10, at 289.

[98] *See In re* Baby M., 537 A.2d 1227, 1252 (N.J. 1988) ("Although the question of best interests of the child is dispositive of the custody issue in a dispute between

A custody determination between biological parents[99] does not ordinarily involve the termination of parental rights. Even a parent whose custodial privileges have been severely curtailed by a previous custody decree retains standing to seek modification of the decree upon proof of changed circumstances.[100] Therefore, in custody cases between parents, such as those pursuant to divorce or separation, where termination of parental rights is not at stake, the welfare of the child provides the standard for the decision.[101]

2. Elements of an Adoption

Adoption involves two separate elements. First, the parental rights of the biological parents must be terminated. Then, an adoption decree is granted according to the welfare of the child.[102] Only after parental rights have been terminated is the focus on the best interests of the child in adoption proceedings consistent with the inquiry in custody proceedings between parents.[103] In fact, when parents are "out of the picture," courts deciding custody cases between nonparents rely on many of the same factors that they rely on when deciding custody disputes between parents.[104]

---

natural parents, it does not govern the question of termination . . . . [T]he mere fact that a child would be better off with one set of parents than with another is an insufficient basis for terminating the natural parent's rights."); *see also In re* Timothy W., 272 Cal. Rptr. 906, 910 (Cal. Ct. App. 1990) ("[C]hildren have no constitutional right to be adopted. Natural parents, on the other hand—even those less suited to parenthood than any available prospective adoptive parent—have a fundamental right to raise their children."); *In re* Clausen, 502 N.W.2d at 668 ("[T]he relationship between natural parents and their children is fundamentally different than that between a child and nonparent custodians."); CLARK, *supra* note 9, § 19.6, at 823.

[99] The term "biological parents" is used throughout this Note in order to provide a clear contrast to the term "adoptive parents." It should be noted that the analysis with regard to "biological parents" actually applies to any legal parent.

[100] *See* CLARK, *supra* note 9, § 19.9.

[101] *See id.* § 19.4.

[102] *See id.* § 20.1, at 850.

All states provide for proceedings independent of the adoption proceedings for termination of parental rights, although in some states the termination of parental rights may alternatively be conducted within the adoption case. *See* GITLIN, *supra* note 75, § 5:03.

[103] *See, e.g., In re* Adoption of Sturgeon, 445 A.2d 1314, 1321 (Pa. Super. Ct. 1982) ("*Once this prior adjudication has been made* and the rights of the natural parents are no more, *the best interests of the child becomes the criterion by which a court must be guided.*") (emphasis added); *In re* Stephanie M., 867 P.2d 706, 718 (Cal. 1994) ("After the termination of reunification services, the parents' interest in the care, custody and companionship of the child are [sic] no longer paramount.").

[104] *See* CLARK, *supra* note 9, § 19.8.

Problems with the application of the UCCJA and the PKPA to adoption proceedings occur in cases pitting parents against nonparents, when the biological parents continue to assert parental rights. In most states, parental rights may be terminated voluntarily (by consent) or involuntarily (upon proof of parental unfitness). Parental unfitness in most states requires a determination of abandonment or neglect.[105] Issues relevant in termination of parental rights proceedings are distinct from those in other custody-related proceedings. The focus of the legal inquiry is on the conduct of the biological parents only. Termination of parental rights involves consideration of a biological parent's consent, abandonment, or neglect; it does not involve consideration of the best interests of the child or an investigation of the suitability of prospective adoptive parents.[106] Unlike other proceedings found to be "custody determinations" within the scope of the UCCJA and the PKPA, decrees of adoption and of termination of parental rights are not perpetually modifiable; neither can be reversed because of changed circumstances. Such final orders are entitled to full faith and credit under the Constitution, regardless of the application of the UCCJA or the PKPA.[107] Ironi-

---

[105] *See id.* § 20.6.

[106] *See, e.g.*, Quilloin v. Walcott, 434 U.S. 246, 255 (1978) ("We have little doubt that the Due Process Clause would be offended '[i]f a State were to attempt to force the breakup of a natural family, over the objections of the parents . . . and for the sole reason that to do so was thought to be in the children's best interest.'" (quoting Smith v. Organization of Foster Families for Equality and Reform, 431 U.S. 816, 862–63 (1977) (Stewart, J., concurring))); *In re* Sanjivini K., 391 N.E.2d 1316, 1320–21 (N.Y. 1979) ("In any event a court may not terminate all parental rights by offering a child for adoption when there has been no parental consent, abandonment, neglect or proven unfitness, even though some might find adoption to be in the child's best interests.").

[107] *See, e.g.*, Byrum v. Hebert, 425 So. 2d 322, 325 (La. Ct. App. 1982); Orme v. Northern Trust Co., 183 N.E.2d 505, 512 (Ill. 1962); Ross v. Pick, 86 A.2d 463, 466–67 (Md. 1952); *see also* CLARK, *supra* note 9, § 20.3, at 869; Schur, *supra* note 45, § 4.02(6), at 4-50.

To the degree that adoption decrees or termination of parental rights orders are entitled to enforcement under the Full Faith and Credit Clause of the Constitution, the modification provisions of the UCCJA and the PKPA are inapplicable. As the Michigan Supreme Court noted in the "Baby Jessica" case:

> [T]here is substantial doubt whether the Iowa decision is the kind of "custody order" that is modifiable at all. When we speak of modifying custody orders, we are ordinarily talking about the typical case of a contest between natural parents. Such orders are at least theoretically perpetually modifiable. Where circumstances change, modification can be made in the child's best interests, because the biological parents have an inherent right to care, custody, and control of the child. That rationale, however, does not apply in a case such as this involving an adoption petition . . . . To say that the order in the instant case is modifiable would have the effect of destabilizing finalized adoptions as well as other final orders.

In re Clausen, 502 N.W.2d 649, 657 n.22 (Mich. 1993).

cally, courts have cited the "finality" of parental rights termination orders in support of the applicability of the UCCJA and PKPA definitions of "custody determination."[108] It is this same characteristic of finality which renders one of the primary concerns of these statutes—the need for statutory limits on the modifiability of child custody decrees[109]—completely moot when the statutes are so applied.

B. *Application to Interstate Adoption Proceedings*

The two primary grounds for jurisdiction under the UCCJA and the PKPA reflect the purposes of the statutes: (1) to identify the jurisdiction in the best position to decide the merits of the case, and (2) to avoid jurisdictional competition among states.[110] The "home state" provision provides a bright-line rule that assigns jurisdiction to the state in which the child has spent a significant period of time immediately preceding the commencement of the custody proceedings on the assumption that any state meeting the "home state" definition should be in the best position to decide the merits of the case. By virtue of being the child's residence, such a state should have access to substantial evidence concerning the welfare of the child.[111] After passage of the PKPA, the "best interest" provision provides a back-up rule that seeks to outline the substantive elements that describe most "home states."[112] The assumption is similar; any state that is significantly connected to a child and has access to substantial evidence about the child should be in a good position to decide questions about the child's welfare.[113] The problem with the application of these provisions to interstate adoption proceedings

---

[108] *See supra* note 63 and accompanying text.

[109] *See supra* notes 17–22 and accompanying text.

[110] *See supra* notes 23, 43; *see also* CLARK, *supra* note 9, § 12.5, at 463, 478–79.

[111] *See, e.g.*, *In re* Clausen, 502 N.W.2d at 674 n.32 (Levin, J., dissenting) ("The home state would ordinarily meet all the criteria of a state qualifying under the alternative; the child and a parent, or the child and a contestant, would necessarily have a significant connection, other than mere physical presence, with the home state, and there would be available substantial evidence concerning the child's present or future care, protection, training, and personal relationships.").

[112] *See id.*

[113] The text of the "best interest" provision makes this assumption explicit; according to the provision, jurisdiction is available in a state where it will be in the *best interest* of the child *because* there is a significant connection with the state and substantial evidence in that state. *See supra* notes 25, 44.

is that these assumptions are inapplicable to cases involving termination of parental rights.

Termination of the rights of biological parents is a threshold issue in adoption cases.[114] In nearly all interstate adoption cases, all the evidence relevant to the termination issue will be located in the state of the biological parents. However, under the UCCJA and the PKPA, contacts and evidence related to a prospective adoptive placement are not distinguished from contacts and evidence related to the termination of parental rights. The state of the adoptive parents and the state of the biological parents are thus equally likely to meet these jurisdictional requirements. Under the UCCJA and the PKPA, the state of the adoptive parents may assert jurisdiction based on a "significant connection" to the child and the adoptive parents, and "substantial evidence" of the child's present or future care in the home of the adoptive parents.[115] However, connections and evidence in the state of the adoptive parents are irrelevant to the issue of the

---

[114] *See supra* notes 102–107 and accompanying text.

Termination of parental rights is also usually the *determinative* issue in adoptive disputes between biological parents and adoptive parents. If parental rights are held to have been validly terminated, a parent may no longer even have standing to challenge the adoption process. *See, e.g.*, *In re* Male Child Born July 15, 1985, 718 P.2d at 664. If parental rights are held not to be validly terminated, prospective adoptive parents are relegated to the role of third-party nonparents. The law in its present form favors parents over nonparents in custody disputes. Many states have adopted a common law rule, commonly known as the "parental right" doctrine, that parents have a right to custody of their children absent proof of unfitness. *See* CLARK, *supra* note 9, § 19.6, at 823–25; Harper v. Landers, 348 S.E.2d 698, 699 (Ga. Ct. App. 1986) ("The law gives parents a *prima facie* right to custody and control of their offspring as against third parties . . . ."); Bubac v. Boston, 600 So. 2d 951, 956 (Miss. 1992) (holding natural parent is entitled to custody, as against third-party grandparent, absent proof of unfitness, abandonment, or immoral conduct adversely affecting children's interests); Wade v. Geren, 743 P.2d 1070, 1075 (Okla. 1987) ("[W]hen the adoption decree had to be set aside [the court] could not simply weigh or compare households; absent a showing of unfitness the father as natural parent would be entitled to custody as against anyone else."); *In re* Michael B., 604 N.E.2d 122, 131–32 (N.Y. 1992) (reasoning that absent a termination of parental rights, an award of custody to a nonparent would leave the child in legal and emotional limbo because the child could not be adopted); *In re* Ronald FF. v. Cindy GG., 511 N.E.2d 75, 77 (N.Y. 1987) ("It has long been recognized that, as between a parent and a third person, parental custody of a child may not be displaced absent grievous cause or necessity."); *see also In re* Baby M., 537 A.2d 1227, 1246–47 (N.J. 1988) (invalidating a surrogacy contract, in part because of policy that "to the extent possible, children should remain with and be brought up by both of their natural parents").

However, the parental right doctrine has received significant criticism. *See, e.g.*, CLARK, *supra* note 9, § 19.6, at 825 ("The parental right doctrine has acquired rigidity from the dubious and amorphous principle that the natural parent has some sort of a constitutional 'right' to the custody of his child. This principle comes dangerously close to treating the child in some sense as the property of his parent . . . ."); *see also infra* notes 148–165 and accompanying text.

[115] For children who are old enough when adopted to qualify for "home state" status,

biological parents' consent, abandonment, or neglect. Thus, since courts may assume jurisdiction based on factors irrelevant to the disposition of the case, the application of the UCCJA and the PKPA to adoption proceedings is not rationally related to the identification of the jurisdiction in the best position to decide the merits of the case. Any correlation between the state in the best position to decide the termination of parental rights issue and the state found to have jurisdiction of the case under the UCCJA and the PKPA is mere coincidence.

III. Substantive Effects of Procedural Arbitrariness: In Most Adoption Cases, Jurisdiction Determines Choice of Law

Repeatedly during testimony in Congressional hearings on the PKPA, witnesses referred to the proper scope of federal legislation.[116] Federal intervention was deemed necessary and proper in declaring jurisdictional rules to settle interstate jurisdictional disputes. However, witnesses were careful to distinguish the procedural provisions of the PKPA from federal legislation of substantive family law. This, they asserted during their testimony, would have been clearly improper.[117] Federal involvement in ad-

---

assertion of jurisdiction by the state of the adoptive parents will prove equally easy, and, under the PKPA, will be exclusive. *See supra* notes 44–46.

[116] *See, e.g., PKA Hearings, supra* note 1, at 20 (statement of Rep. Duncan) ("It is a complex and delicate question of conflict of laws . . . . Should the Federal Government involve itself in a situation in which a parent is punished for moving his own child without harm, across a State line? How far should Government go into social policy? Should Government intervene in a private marital problem? When does it become a public problem?").

[117] *See PKA Hearings, supra* note 1, at 146 (Professor Russell M. Coombs, Associate Professor, Rutgers University School of Law) ("The provisions of S. 105 . . . display a clear recognition of the difference between questions of conflicts of jurisdiction and full faith and credit on the one hand—questions that are suitable for Federal legislation—and, on the other hand, questions of substantive family law and details of procedure and practice."); *Criminal Reform Hearings, supra* note 36, at 10,627 (Professor Russell M. Coombs, Associate Professor, Rutgers University School of Law) ("When disputes between parents over their children must be resolved, the States must carry the primary burden. This bill carefully observes that principle. It is confined to dealing with conflicts between States, not conflicts between parents."); *Parental Kidnaping Hearings, supra* note 12, at 140 (Professor Russell M. Coombs, Associate Professor, Rutgers University School of Law) (criticizing alternative legislation because of "the failure of these bills to preclude federal courts from undertaking the clearly inappropriate task of developing a federal substantive law of custody to be applied in selective 'enforcement' of custody orders.").

Proposed language that would have allowed states to reverse the custody decrees of other states upon finding a decree "punitive" or "against public policy" was repeatedly criticized and finally struck from the bill for showing "insufficient regard for the proper

ministering the PKPA has been minimized through both Congressional and judicial action. Congress soundly rejected an alternative parental kidnaping bill that explicitly granted jurisdiction to federal district courts to issue warrants upon the violation of state custody decrees.[118] The Supreme Court ended the involvement of the federal courts in interpreting the PKPA in *Thompson v. Thompson*, holding that the PKPA was only procedural in function, and thus created no cause of action justiciable in federal courts.[119]

The analysis in Part I and Part II has illustrated the arbitrariness of jurisdictional outcomes in adoption cases under the UCCJA and the PKPA. The PKPA policy favoring the clear directive of "home state" jurisdiction is often thwarted. The policy of identification of the best forum is not promoted. In termination of parental rights cases, modifiability is not an issue, and thus the PKPA policy to curb modifiability of custody decrees is not furthered.

The UCCJA and PKPA do serve other purposes. Provisions in both acts prevent the assertion of jurisdiction by a state if the same matter is pending in another state.[120] When an interstate adoption proceeding has already been filed in a state meeting the jurisdictional requirements of the acts, the "simultaneous proceedings" provision does prevent some measure of further jurisdictional competition. If the only negative effects of arbitrary jurisdictional determinations under the Federal Act were merely procedural, perhaps the PKPA would be substantively innocuous. In such a case, the application of the UCCJA and the

---

relationships among states and between the federal government and the states." *Parental Kidnaping Hearings*, *supra* note 12, at 134 (statement of Russell M. Coombs, Rutgers University School of Law); *see also PKA Hearings*, *supra* note 1, at 132 (statement of Nancy Lynn Hiestand, Juvenile Justice Clinic, Georgetown University Law Center).

[118] H.R. 9913, 96th Cong., 2d Sess. (1980). *See Parental Kidnaping Hearings*, *supra* note 12, at 11 (statement of Rep. Sensenbrenner) ("Your [addressing the sponsor] alternative is an interesting one which may expand the jurisdiction of the Federal district courts into family law perhaps for the first time in history . . . .").

[119] 484 U.S. 174, 187 (1988). The Ninth Circuit opinion in *Thompson* appealed to the policies of the domestic relations exception to diversity jurisdiction when holding that the PKPA does not create a federal cause of action:

> The PKPA is so structured that in a type of case likely to arise frequently, a federal court deciding which of two conflicting state court decrees is valid under the PKPA could not avoid becoming involved in the merits of the underlying dispute.

Thompson v. Thompson, 798 F.2d 1547 (9th Cir. 1986), *aff'd*, 484 U.S. 174 (1988).

[120] *See* UCCJA § 6; PKPA § 8(g).

PKPA to adoption proceedings would be justifiable, despite the arbitrariness of their ill-suited provisions.

However, despite the insistence of PKPA proponents that the statute would not interfere with the substantive family law of the states,[121] jurisdictional determinations undeniably have substantive effects. One of the primary evils addressed by the PKPA, in fact, was the incentive for parents to forum shop—i.e., remove children to other jurisdictions in order to seek more favorable results on the merits.[122] With regard to custody disputes between parents, a change of jurisdiction rarely means a change of the applicable legal standard: the best interest of the child.[123] Varying outcomes of such cases in different states depend on different applications of the same standard.[124] With regard to custody disputes in the adoption context, the state adoption laws themselves vary widely in numerous aspects.[125] Thus, in addition to differences in judicial applications of uniform standards, jurisdictional determinations in adoption cases often create conflict-of-laws problems.[126] Despite recent trends applying the "most significant contacts" test to solve conflict-of-laws problems in other areas of law, in adoption cases, the Restatement (Second) of Conflict

---

[121] *See supra* note 117.

[122] *See supra* notes 16–22, 37–42 and accompanying text; *see also* PKPA § 7(a)(3) ("[Parents] involved in [custody] disputes . . . frequently resort to the seizure, restraint, concealment, and interstate transportation of children, the disregard of court orders, excessive relitigation of cases, obtaining of conflicting orders by the courts of various jurisdictions, and interstate travel and communication that is . . . expensive and time consuming . . . .").

[123] *See supra* note 101.

[124] Granted, the best interest standard is inherently vague and has been criticized for its susceptibility to the discretion of trial courts. *See* CLARK, *supra* note 9, § 19.4, at 797–98.

[125] For example, standards of parental consent to adoption have historically varied greatly from state to state. *See* Tracy A. Bateman, Annotation, *Validity of Birth Parent's "Blanket" Consent to Adoption Which Fails to Identify Adoptive Parents*, 15 A.L.R.5th 1 (1993); Gary D. Spivey, Annotation, *Comment Note—Right of Natural Parent to Withdraw Valid Consent to Adoption of Child*, 74 A.L.R.3d 421 (1976); T.C. Williams, Annotation, *Annulment or Vacation of Adoption Decree by Adopting Parent or Natural Parent Consenting to Adoption*, 2 A.L.R.2d 887 (1948).

Different legal standards can create incentives for forum shopping. For example, some states have been labeled "baby markets" because of a lack of state law restrictions on adoptions. *See* Joan M. Cheever, *Lone Star State Legislators Prepare to Apply the Brakes on the So-Called Baby Train*, NAT'L L.J., Aug. 17, 1992, at 8; Gordon Oliver, *Oregon's Adoption Regulations May Be Loosest in the Nation*, PORTLAND OREGONIAN, May 16, 1993, at A1.

[126] *See, e.g.*, Stubbs v. Weathersby, 869 P.2d 893, 898 (Or. Ct. App. 1994) (applying Oregon law of consent as opposed to Washington law of consent); *In re* Adoption of Child by T.W.C., 636 A.2d 1083, 1090 (N.J. Super. Ct. App. Div. 1994) (applying New Jersey law as opposed to New York law).

of Laws directs that the law of the forum be applied.[127] Some courts have gone so far as to state that in adoption cases, there is no distinction between choice of jurisdiction and choice of law.[128] The determinative issue in most interstate adoption cases between adoptive parents and biological parents is the validity of the termination of parental rights. This question almost always depends on evidence in the state of the biological parents.[129] The PKPA not only allows the adoptive parents' state to assert jurisdiction over such a case, resulting in the application of forum law,[130] but the PKPA also requires the biological parents' state to enforce the resulting order.[131] Thus, under the PKPA, the parental rights of biological parents may be terminated according to the law of the adoptive parents' state. This is the case even if all conduct relating to the termination occurred in the biological parents' state; this is also the case even if application of the law of the biological parents' state would have yielded the opposite result. Therefore, the jurisdictional provisions of the PKPA threaten the ability of states to regulate parental conduct within their borders, a sphere of family law traditionally free from federal interference.[132]

---

[127] RESTATEMENT (SECOND) CONFLICT OF LAWS § 289 (1969).

[128] *See Stubbs*, 869 P.2d at 899 (interpreting Restatement to suggest that "once jurisdiction has been decided in an adoption case, there is no choice of law problem [because] local law will always apply in adoption cases"); *In re* Adoption of Child by T.W.C., 636 A.2d at 1090; *In re* Adoption of C.L.W., 467 So. 2d 1106, 1111 (Fla. Dist. Ct. App. 1985).

Jurisdiction may also determine the substantive result on the merits due to the burden of litigation expenses. *See In re* B.B.R., 566 A.2d 1032, 1042 (D.C. 1989) (Schwelb, J., concurring) ("In theory, the choice is simply which of two impartial judicial systems will decide the case. In the real world, especially for a non-affluent party, the opportunity to litigate the case meaningfully may be at stake.").

[129] *See supra* note 114.

[130] *See, e.g., Stubbs*, 869 P.2d at 898 (applying Oregon law of consent, instead of Washington's more rigorous standard of consent, to the case of a child born in Washington to a Washington resident).

[131] *See supra* note 44, 45.

[132] For example, consider a case in which the biological parents seek to revoke consent to termination of parental rights five days after birth. If the biological parents' state allowed revocation up to seven days after birth, but the prospective adoptive parents' state allowed revocation of consent only within 48 hours after birth, the outcome of the case would depend on the choice of law issue.

This problem is unique to the PKPA; the Supremacy Clause is not at issue when the UCCJA conflicts with other state statutes. *See* Schur, *supra* note 45, § 4.02(7), at 4-53 n.89.

IV. Potential Amendments and Recommendations: Jurisdictional Rules Should Be Squared with Interests Unique to Adoption Cases

The above analysis illustrates that the textual provisions of the UCCJA and the PKPA are ill-tailored to settle jurisdictional disputes in adoption cases, that the application of these provisions to interstate adoption cases is inconsistent with the general purposes of the UCCJA and the PKPA, and that these procedural provisions have a detrimental effect on the ability of states to regulate the termination of parental rights. Yet even after the problem of the application of the UCCJA and the PKPA to interstate adoption cases is recognized, the challenge of developing solutions remains.

A. *Definitional Amendments*

At first glance, several textual amendments to the UCCJA and PKPA present themselves as potentially quick and easy solutions. The definition of "custody proceeding" ("custody determination" in the PKPA) could be amended to specifically exclude adoption proceedings.[133] Such an amendment would eliminate the substantive problems caused by the application of the UCCJA and the PKPA to adoption cases, and the Full Faith and Credit Clause would still be available to ensure interstate enforcement of judgments terminating parental rights. However, such an amendment would leave no mechanism to resolve jurisdictional disputes arising from interstate adoptions. Clear jurisdictional rules in interstate adoption cases are needed both to identify the proper initial jurisdiction and to prevent simultaneous adjudications.[134] However, any such rules should recognize the distinctive nature of adoption cases so that resulting jurisdictional determinations are neither arbitrary nor adverse to the purposes of the statutory scheme. Ultimately, there must be a

---

[133] Some states have enacted a variation to the UCCJA definition of "custody proceeding" that excludes adoptions. *See supra* note 56.

[134] Although illegal abductions are rarer in interstate adoption cases than in interstate custody cases between parents, they do occur. *See, e.g.*, Owens *ex rel.* Mosley v. Huffman, 481 So. 2d 231, 239 (Miss. 1985) ("While the two Acts are directed to child custody disputes which occur *between parents*, the conduct of the Mosleys in this case is a striking example of the evil PKPA and UCCJA are designed to discourage and prevent.").

better answer than defining adoptions out of the legislation altogether.

Alternatively, the definition of "home state" could be amended so that it would not deny "home state" status to the vast majority of infants involved in interstate adoption disputes. For example, a few states have enacted a variation of the UCCJA "home state" definition that allows jurisdiction in the state where the child has lived "a majority of the time" since birth if the child is less than six months old.[135] Under this definition, most infants involved in interstate adoptions would have an identifiable "home state." Such an amendment would prove consistent with the UCCJA and PKPA policy to avoid jurisdictional disputes and would end the race-to-the-courthouse jurisdiction that results from application of the present statutory language. However, this amendment to the "home state" definition would not end incentives for strategic filings. For example, biological parents who did not file an action challenging the adoption immediately would have an incentive to wait to file until after the child turns six months old but before the child had lived for six months in the adoptive placement. The biological parents would thereby avoid both conditions of the "home state" test.[136]

Even if a "home state" could be identified for every child subject to an interstate adoption, the jurisdictional results in

---

[135] *See supra* note 82. The amended "home state" definition would read as follows: "[H]ome state" means the state in which the child immediately preceding the time involved lived with his parents, a parent, or a person acting as a parent, for at least 6 consecutive months, and in the case of a child less than 6 months old a state in which the child lived a majority of the time from birth with any of the persons mentioned. Periods of temporary absence of any of the named persons are counted as part of the 6 month or other period.

[136] The facts of the "Baby Jessica" case can be used to illustrate. *See In re* Clausen, 502 N.W.2d 649 (Mich. 1993). After birth, Jessica spent two weeks in Iowa with state-appointed caregivers before being placed with the DeBoers in Michigan. Under a "majority of the time" standard, Michigan would have become the child's "home state" after she had lived in Michigan for more than two weeks. However, when the child turned six months old, she could no longer qualify for "home state" status under the "majority of the time" standard (because it would apply only to "the case of a child less than six months old"). At that time, Michigan would not have met the other prong of the "home state" definition either (because it requires residence for "at least six consecutive months," and she would have only resided in Michigan for five months, two weeks). Thus, the only opportunities for a biological parent to commence proceedings to challenge an adoption when the state of the adoptive placement would not be the "home state" of the child, and thus have exclusive jurisdiction over the case under the PKPA, would be the time period immediately after the adoptive placement and the time period immediately after the child turns six months old. Within these windows, either state could likely assert "best interest" jurisdiction, and the first state to do so would properly have control of the case.

adoption cases under the UCCJA and the PKPA would continue to be only arbitrarily related to the identification of the best forum. In custody disputes following a divorce or separation, there is a rational relationship between the identification of the child's "home state" and the identification of the forum in the best position to decide the determinative issue in the case, the best interests of the child.[137] In an adoptive dispute, the determinative issue in the case is usually the validity of the termination of parental rights.[138] There is no rational relationship between a standard based on the amount of time a child is in a jurisdiction and the identification of the state in the best position to hear evidence about the conduct of biological parents. In fact, in adoption cases, since the state that has the best access to evidence concerning the validity of the termination will almost always be the state of the biological parents, any relationship between a "home state" provision based on a "majority of time" standard and the identification of the state of the biological parents would likely be inverse. The "majority of time" amendment to the "home state" definition is thus unsatisfactory.

B. *Uniform Adoption Legislation*

One way to solve the problem of arbitrariness in the determination of jurisdiction is to negate the substantive effects of the determination. The National Conference of Commissioners on Uniform State Laws (NCCUSL) has recently approved a new Uniform Adoption Act.[139] To the degree that standards of consent to adoption, for example, are the same from state to state, the choice of law consequences of an assignment of jurisdiction will be minimized.

Uniform legislation can prove highly successful in easing conflict-of-law problems. Yet, as has been the case with the UCCJA, the solution is rarely a complete one. All states must enact the

---

[137] *See supra* notes 96–101 and accompanying text.

[138] *See supra* note 114.

[139] The new UAA was approved at the August 1994 annual meeting of the NCCUSL in Chicago. *See Legislative Update: Uniform Adoption Act Is Approved*, FAIR$HARE, Nov. 1994, at 26. Recommendation of the Act was on the agenda for the 1995 Midyear Meeting of the A.B.A. *See* Vicki S. Porter, *Agenda for the ABA House of Delegates at the 1995 Midyear Meeting*, 24 COLO. LAW. 19, 20 (1995). For a general discussion of the legislation, focusing on the issues of parental consent, race-based matching, and open adoption policies, *see* Salmine, *supra* note 73.

legislation in order for the law to be truly uniform.[140] A conflict will still arise in an interstate adoption case between a state that has adopted the Uniform Act and a state that has not. Secondly, states are free to enact variations of recommended uniform legislation. Variations of key provisions can undermine uniformity and create conflicts, even between states that both have substantially enacted the uniform law. Even if the same language has been enacted by states involved in interstate adoptions, differing judicial interpretations of key provisions in different states can also create conflicts. Since federal legislation preempts conflicting state law under the Supremacy Clause, the PKPA will require Uniform Act states to enforce decisions not in conformity with the uniform adoption legislation, so long as the state rendering the decision meets the jurisdictional conditions of the PKPA.[141] Uniform adoption legislation therefore may be highly desirable, but does not foreclose the need for further action to make the UCCJA and the PKPA more compatible with adoption cases.

C. *Jurisdictional Preference for Biological Parents*

The interests addressed by the UCCJA and the PKPA and the interests relevant in most interstate adoption proceedings are plainly incongruous. The UCCJA and the PKPA were intended primarily to address custody disputes between parents, in which the best interests of the child are determinative.[142] The jurisdictional provisions of the UCCJA and PKPA thus focus on the child—the child's physical location, the child's connections, the child's present and future care. However, the central issue in most custody disputes between adoptive parents and biological parents is the validity of the termination of parental rights, in

---

[140] Only five states enacted the previous version of the Uniform Adoption Act, promulgated in 1969 and amended in 1971. *See* UNIFORM ADOPTION ACT, 9 U.L.A. 11 (1988). Numerous attempts have been made to promulgate Uniform Acts and Model Acts concerning adoption law. The 1988 DRAFT A.B.A. MODEL STATE ADOPTION ACT, which explicitly applied the UCCJA to adoption proceedings, is one of the most recently failed efforts. *See* Schur, *supra* note 45, § 4.02(7), at 4-58, § 4.07(4), at 4-127.

[141] *See In re* B.B.R., 566 A.2d 1032, 1036 n.10 (D.C. 1989); *In re* Adoption of Zachariah K., 8 Cal. Rptr. 2d 423, 429 (Cal. Ct. App. 1992); *see also* CLARK, *supra* note 9, § 12.5, at 477.

[142] *See supra* notes 13, 96, 101.

which the best interests of the child are largely irrelevant.[143] If the UCCJA and the PKPA are to be applied to interstate adoption cases, this incongruity should be corrected. The jurisdictional provisions of the UCCJA and the PKPA could be reformulated in order to distinguish between connections and evidence relevant to termination of parental rights and connections and evidence relevant to post-termination adoptive placement. The "home state" provision could be redefined to add an additional condition specifying the state of the biological parents as the "home state" in interstate adoption cases.[144] The "best interest" provision could be amended to limit its application in interstate adoption cases to significant connections with the biological parents and substantial evidence about the validity of the termination.[145] Regardless of the form of such changes, the purpose would be to establish a bright-line rule: in cases where termination of

---

[143] *See supra* note 114 and accompanying text.

[144] This proposed revision to the "home state" definition could take a form similar to the following:

> "[H]ome state" is the state where one of the following conditions is met:
>
> (a) such state is (1) the state of residence of the parent(s) in proceedings in which the validity of the termination of the parental rights of those parent(s) is contested, and (2) the state of birth of the child at issue in the proceeding; or
>
> (b) (1) it appears that no state is the "home state" under paragraph (a), and (2) the child immediately preceding the time involved lived with his parents, a parent, or a person acting as a parent, in such state for at least 6 consecutive months, and in the case of a child less than 6 months old the child lived from birth in such state with any of the persons mentioned. Periods of temporary absence of any of the named persons are counted as part of the 6-month or other period.

[145] This proposed revision of the "best interest" provision could take a form similar to the following:

> A court has jurisdiction to make a child custody determination by initial or modification decree if one of the following conditions is met:
>
> (a) . . . .
>
> (b) (1) it appears that no other state would have jurisdiction under paragraph (a), and (2) it is in the best interest of the child that a court of such state assume jurisdiction because:
>
> (A) (i) in a case involving the validity of a termination of parental rights, the parent(s) whose rights are at issue have a significant connection with such state other than mere physical presence in such state, or (ii) (I) it appears that no state satisfies the requirements of subparagraph (A)(i), and (II) the child and his parents, or the child and at least one contestant, have a significant connection with such state other than mere physical presence in such state, and
>
> (B) (i) in a case involving the validity of a termination of parental rights, there is available in such state substantial evidence concerning the conduct of the parent(s) whose parental rights are at issue, or (ii) (I) it appears that no state satisfies the requirements of subparagraph (B)(i), and (II) there is available in such state substantial evidence concerning the child's present or future care, protection, training, and personal relationships.

parental rights is an issue, only the state of the biological parents should assert jurisdiction. This preference for biological parents over prospective adoptive parents is consistent with current law, which recognizes the special protections afforded parental rights and generally does not recognize children's rights independent of their parents' interests.[146] Such a jurisdictional rule would also be consistent with the statutory purposes of the UCCJA and the PKPA to avoid jurisdictional competition and to identify the jurisdiction in the best position to decide the merits of the case.[147]

---

[146] *See supra* notes 97–98 and accompanying text.

[147] As discussed earlier, *supra* text accompanying notes 69–72, Professor Clark in his family law treatise also recognizes the need to establish a bright-line jurisdictional rule in adoption cases. However, Professor Clark concludes that the rule should favor jurisdiction in the state of the adoptive parents. *See* CLARK, *supra* note 9, § 20.3, at 873. This result seems odd because his analysis, too, begins with the premise that "[t]he purpose of jurisdictional rules, in this field at least, is to authorize adoption decrees by the court best able to evaluate the conflicting claims." His analysis continues:

> In practice this means the court which can investigate and judge the qualifications of the prospective adoptive parents. The rights of the natural parents are terminated in adoption cases either by their consent, or involuntarily upon proof of abandonment, neglect, non-support or the like. The only issues raised with respect to the natural parents are therefore whether the consent is genuine, or whether the alleged abandonment or neglect did occur. These resemble the issues in the ordinary transitory lawsuit and there is thus no need for any requirement of domicile or residence on the part of the natural parents.

CLARK, *supra* note 9, § 20.3, at 872.

In this passage, Professor Clark makes the same error of assuming the parity of interests of adoptive parents and biological parents that is made by the UCCJA and the PKPA themselves. His analysis correctly identifies the differences in the types of issues relevant to each party's custody claim, but then directly compares the nature of those issues to determine for which issues the proximity of the court is more important. True, a court might find it more difficult to investigate the background of an out-of-state adoptive family than to discern the facts relevant to the consent or neglect of out-of-state biological parents. However, no court hearing an interstate adoption dispute is faced with this choice of lesser evils, because these issues are not considered simultaneously. The fitness of a prospective adoptive family is absolutely irrelevant in an adoption case until parental rights have been validly terminated. In these cases, the "conflicting claims" do not pit the biological parents' claim of fitness versus the adoptive parents' claim of fitness. Under current law, the conflict is between the biological parents' claim that their parental rights have not been validly terminated and the adoptive parents' claim that the biological parents' rights have been validly terminated. *See supra* note 114. Thus, since the termination of the biological parents' rights is at the heart of the claims of both parties, the "court best able to evaluate the conflicting claims" must surely be the state where evidence relevant to the alleged termination is located, which will usually be the state of the biological parents.

It is also interesting to note that Professor Clark's additional observation that the UCCJA and the PKPA implement his analysis, *see supra* text accompanying note 71, also seems to have proven false. Since, as illustrated in part I and part II above, the assignment of jurisdiction in interstate adoption cases is quite arbitrary, these statutes presently fail to favor jurisdiction in either the adoptive parents' state or the biological parents' state.

D. *A Final Thought: Recognition of the Independent Interests of Children*

Establishing a bright-line rule vesting jurisdiction in interstate adoption cases in the state of the biological parents seems necessary to make application of the UCCJA and the PKPA to adoption proceedings consistent with their statutory purposes under present law. However, the substance of the present law of the parental right doctrine, which does not acknowledge the interests of children independent of the interests of their biological parents,[148] is not altogether palatable. As outlined above, one way to correct the incongruity between the interests embraced by the UCCJA and the PKPA and the unique interests involved in interstate adoption proceedings is to amend the UCCJA and the PKPA. Though not squarely within the scope of this Note, there is an alternative solution. The interests at stake in adoption cases could be redefined to recognize the independent interests of children, which would more closely reflect the interests embraced by the UCCJA and the PKPA. In other words, custody disputes between biological parents and prospective adoptive parents could be adjudicated according to standards presently applicable to custody disputes between parents.

The constitutional liberty interests of parents must surely be protected, but those interests are not absolute.[149] Lawmakers and judges should ask themselves why the interests of children are given no independent significance in adoption cases until after parental rights have been terminated. In cases in which the child involved has never met—much less lived with—the biological parents, it seems little more than a legal fiction to assert that the interests of the child are in harmony with the interests of the biological parents.[150] Whether constitutionally recognized or not, children have identifiable interests that are distinct from the

---

[148] *See supra* notes 97–98, 114 and accompanying text.

[149] Even Supreme Court cases focusing on the due process rights of biological parents have allowed termination of those rights upon proof of parental unfitness. *See supra* notes 93–96.

[150] *Cf. In re* Clausen, 502 N.W.2d 649, 668–69 (Mich. 1993) (Levin, J., dissenting) ("[T]his is not a lawsuit concerning the ownership, the legal title to a bale of hay . . . . Jessica DeBoer . . . will now be told . . . that she is not Jessie, that the DeBoers are not Mommy and Daddy, that her name is Anna Lee Schmidt, and that the Schmidts, whom she has never met, are Mommy and Daddy."); Bennett v. Jeffreys, 356 N.E.2d 277, 281 (N.Y. 1976) ("[A] child is a person, and not a subperson over whom the parent has an absolute possessory interest.").

interests of their biological parents.[151] Both the UCCJA and the PKPA recognize a government interest in a "greater stability of home environment and of secure family relationships for the child."[152] Repeatedly, witnesses testifying at congressional hearings on the PKPA focused on the harmful effects of disrupting a child's stable home environment.[153] In testimony on the question of whether the Act should cover cases of child-snatching before the entrance of a custody decree, a witness observed:

> I think it is grossly unfair to assume that because a child was taken in violation of a court order, they are somehow more traumatized than a child who by happenstance did not happen to be protected by such a court order at that time.[154]

Indeed, harm to a child is not dependent on such legal distinctions. It must certainly be traumatic for a child to be snatched by a non-custodial parent in violation of a court order from the home of a custodial parent. It certainly could not be any less traumatic for a child snatched pursuant to a court order from the home of would-be adoptive parents to be placed in the unfamiliar home of biological parents.[155] Excerpts from the testimony in

---

[151] *See* GOLDSTEIN, *supra* note 6, at 31–49.

[152] UCCJA § 1(a)(4); PKPA § 7(C)(4).

[153] *See Parental Kidnaping Hearings*, *supra* note 12, at 104 (testimony of Dr. Doris Jonas Freed, Chairperson, A.B.A. Custody Committee); *id.* at 144–45 (statement of Lee H. Haller, M.D., child, adolescent, and forensic psychiatrist); *PKA Hearings*, *supra* note 1, at 65 (testimony of Andrew Yankwitt, counsel, Citizens League on Custody Kidnapping); *id.* at 105–06 (testimony of Dr. Jeanette I. Minkoff, probation-family services coordinator, Monroe County, N.Y.); *Legislation to Revise and Recodify Federal Criminal Laws, Part 2: Hearings on H.R. 6869 Before the Subcomm. on Criminal Justice of the House Comm. on the Judiciary*, 95th Cong., 1st & 2d Sess. 1016 (1978) (statement of Ramona Powell, Georgetown Juvenile Justice Clinic); *cf.* UCCJA Prefatory Note, 9 U.L.A. at 116:

> It does not require an expert in the behavioral sciences to know that a child, especially during his early years and the years of growth, needs security and stability of environment and a continuity of affection. A child who has never been given the chance to develop a sense of belonging and whose personal attachments when beginning to form are cruelly disrupted, may well be crippled for life, to his own lasting detriment and the detriment of society.

[154] *PKA Hearings*, *supra* note 1, at 75 (testimony of Rae Gummel, vice president, Children's Rights, Inc.).

[155] *Cf., e.g.*, Wade v. Geren, 743 P.2d 1070, 1075 (Okla. 1987) (taking nine-year-old child away from only home she had ever known and awarding custody to natural father she had never met).

For an interesting, though somewhat heartless, opposing view, consider Faupel, *supra* note 10, at 293:

> The possibility of trauma is not enough to defeat constitutional mandates. In fact, many reasonable decisions and circumstances cause trauma to family members. There is trauma when a parent is sent overseas for military duty,

Congressional hearings on the PKPA concerning the evils of child-snatching certainly seem applicable to children like Baby Richard and Baby Jessica as well:

> Children develop coping mechanisms and attempt to find pleasure and security in a familiar environment, a special toy, peers, a teacher, a friend, etc. and when a child is snatched from his familiar environment all of a sudden even these special little things are out of his grasp. The child is stripped of everything he has to identify with, including the other parent. The child or children are now forced into a new environment, new school, and new home with nothing familiar but the clothes on his back.[156]

> [T]he trauma they have suffered has had a profound psychological effect. Although the specific symptoms vary with the age and personality of the children, almost invariably they return as troubled youngsters . . . . [T]he children may become fearful, anxious, depressed and withdraw into themselves . . . . The move or moves may cause them to be unable to integrate into new settings. One manifestation of this may well be marked difficulties in concentration in school, such that they fall behind academically as well as emotionally.[157]

A substantive reformulation of adoption law and a complete examination of the constitutional implications of such reforms is well beyond the scope of the UCCJA and the PKPA. However, in light of the problems caused by application of the UCCJA and the PKPA to interstate adoption cases, it seems essential not only to consider how to change the PKPA and the UCCJA to conform with existing adoption law, but also to consider how existing adoption law might be adjusted to better serve the interests advanced by the UCCJA and the PKPA. These approaches to reform are not by any means mutually exclusive. The legislative proposal of a bright-line jurisdictional rule favoring the state of the biological parents would surely highlight the differences between adoption cases and other custody disputes, which could

---

hospitalized for months with a debilitating disease, or incarcerated. The courts and their many litigants are no strangers to trauma. In the end, however, what the law balances is the trauma to one individual as a result of the application of the law in an individual case against the threat of trauma to an entire society if that society becomes lawless.

[156] *PKA Hearings*, *supra* note 1, at 116 (statement of Dr. Jeanette I. Minkoff, probation-family services coordinator, Monroe County, N.Y.).

[157] *Parental Kidnaping Hearings*, *supra* note 12, at 144–45 (statement of Lee H. Haller, M.D., child, adolescent, and forensic psychiatrist).

push lawmakers to confront the basis for the distinction. Instead of amending the UCCJA or the PKPA to instruct courts to ignore the interests of the interstate child in adoption cases, legislatures may be motivated to find a way, consistent with constitutional considerations, to recognize that the interests of a child caught in the middle of an adoptive dispute may be independent of the interests of the biological parent.

There are at least two possible approaches to the recognition of the independent interests of such children. First, the standard for termination of parental rights could be shifted from a parental unfitness standard to a best interests standard. Statutes in a few states presently authorize termination of parental rights if in the best interests of the child, but in practice these statutes do not operate significantly differently from a standard of parental unfitness.[158] A standard that allowed parental rights to be terminated because the child would be "better off" in another family would raise multiple constitutional concerns. First, such a standard seems contrary to Supreme Court rulings recognizing that biological parents have a due process right to a full and fair opportunity to litigate the question of parental unfitness prior to any termination of parental rights.[159] In addition, under a best interests standard, poor families that could not afford quality legal representation would provide easy targets for wealthy couples unable to adopt through ordinary channels. Thus, application of a best interests standard would implicate equal protection issues for groups such as minorities, young people, and the disabled who could be disproportionately targeted for such termination decrees.[160]

Second, even if termination of parental rights standards were left undisturbed, the best interests standard could be applied to

---

[158] *See* CLARK, *supra* note 9, § 20.6, at 905. There is no necessary inconsistency between a best interests standard and a parental unfitness standard. Many courts have simply interpreted the best interests standard in this context by establishing a conclusive presumption that a child's best interests will be served by custody with the natural parents, absent proof of parental unfitness. *See, e.g., In re* Sanjivini K., 391 N.E.2d 1316, 1321 (N.Y. 1979); *In re* New England Home for Little Wanderers, 328 N.E.2d 854, 860–61 (Mass. 1975). Under such an interpretation, the best interests standard does not recognize the interests of children independent of the interests of their parents and so functionally remains a parental unfitness standard.

[159] *See supra* notes 96–98 and accompanying text.

[160] *See In re* C.C.R.S., 892 P.2d 246, 262 (Colo. 1995) (Lohr, J., dissenting) ("The best interests test can all too easily disadvantage the poor, the uneducated, and the otherwise disadvantaged individuals."); *see also* Faupel, *supra* note 10, at 310, 313 (noting the potential for "social engineering" under a best interests standard for termination of parental rights).

the issue of custody in disputes between biological parents and would-be adoptive parents.[161] This alternative would shift the focus of inquiry in such cases from the rights of the parents to the interests of the child.[162] Recognition of the interests of the child in custody disputes between biological parents and non-biological, custodial parents would have the value of protecting a child's interest in stability and continuity and would likely prevent much of the trauma that children experience when uprooted from their homes in their formative years.[163]

---

[161] For a discussion of the parental rights doctrine, which presently governs such questions in most states, *see supra* note 114.

[162] In practice, a shift from the interests of the parent to the interests of the child is also a shift from the interests of parents to the interests of non-parent custodians. There can never be true "parity" in a custody dispute between biological parents and third-party custodians. Just as a "parental fitness" standard favors most non-custodial biological parents, a "best interests" standard would generally favor the parties in possession of the child, as the majority of best interests standards tend to favor the status quo. This "status quo bias" emerges from the judicial recognition of a child's interest in stability and continuity. *See, e.g.*, *In re* Sturgeon, 445 A.2d 1314, 1322 (Pa. Super. Ct. 1982) ("We do know that David's future under present conditions is best served by the stability of his present home, and that must be the guide for this court."); *In re* Aldridge, 841 S.W.2d 793, 803 (Mo. 1992) ("[E]very day a child in its formative years is left in a stable parent-child relationship, natural or foster, the greater the potential for harm to the emotional well-being of the child should it be necessary to order a change of custody . . . . Chris' best interest is served by leaving him where he is."); Burchard v. Garay, 724 P.2d 486, 490–91 (Cal. 1986) (noting that a child's need for stability and continuity will often dictate that maintenance of the current arrangement is in the child's best interests); *see also* Faupel, *supra* note 10, at 285.

[163] Though certainly a minority position, there is both statutory and judicial precedent for the application of the best interests standard in these situations. *See In re* C.C.R.S., 892 P.2d at 248, 256–58 (applying best interests standard to determine custody issue between natural parents and "psychological parents," even after valid revocation of consent to adoption by biological mother); Owens *ex rel.* Mosley v. Huffman, 481 So. 2d 231, 244 (Miss. 1985) (upholding lower court's best interests determination as "in accord with the spirit and purpose of the PKPA and UCCJA"); Lemley v. Barr, 343 S.E.2d 101, 109 (W. Va. 1986) (remanding for best interests determination before considering transfer of custody to the biological parents, "who are complete strangers to" the child).

In the appeals court decision in the Baby Richard case, *see supra* text accompanying notes 2–4, the majority applied a form of the best interests standard when it creatively interpreted federal and state legislation to construct a bright-line rule preventing the removal of an infant from an adoptive placement after 18 months. *In re* Doe, 627 N.E.2d 648, 653 (Ill. App. Ct. 1993). After the Illinois Supreme Court reversed that decision in *In re* Doe, 638 N.E.2d 181 (Ill. 1994), Governor Jim Edgar called a special session of the state legislature to consider a bill that would impose a best interests standard:

> "I think the court changed the rules. These parents had adopted a child, and I think most everyone feels when you adopt the child, the child is then yours . . . . The court here has overturned two lower decisions . . . and they're yanking that child away from the only parents that child has ever known. To me, that's what's outrageous."

Edgar Calls for Special Session on "Baby Richard" Legislation, CHI. DAILY L. BULL., July 1, 1994, at 1 (quoting news conference of Governor Edgar).

The bill, providing for a best interests hearing in the event that an order for adoption

This proposal does have disadvantages. Opponents have argued that recognition of the independent interests of children would only provide a mechanism for disappointed adoptive parents to use the child to circumvent adoption laws.[164] In addition, would-be adoptive parents, even if awarded custody of the child, would not be able to adopt the child if parental rights were not terminated. Such a custody determination could leave the child in a state of "legal limbo" that might itself be contrary to the child's interest in stability.[165] No solution will be perfect, but imperfection should not be allowed to forestall efforts to implement legislation that will best serve the interests of children involved in adoptive custody disputes.

CONCLUSION

The Uniform Child Custody Jurisdiction Act and the Parental Kidnaping Prevention Act were legislative responses to the state of jurisdictional chaos that resulted largely from the modifiability of child custody decrees. When interstate adoptions led to similar jurisdictional disputes, the UCCJA and the PKPA seemed to provide easy solutions. However, application of these statutes has succeeded only in making interstate adoptions more complicated. Definitions are not suited to adoption cases. Jurisdictional rules produce results contrary to the purposes of the statutes, one result being a federal intrusion into areas historically reserved for state control. It is quite possible to amend the statutory language to make the UCCJA and the PKPA consistent with

---

is vacated, was enacted as ILL. ANN. STAT. ch. 750, para. 50/20b (Smith-Hurd 1995). The statute was made applicable to "all cases pending" on July 3, 1994, which apparently was intended to include the Baby Richard case itself.

In its opinion on the habeas petition for return of the child, decided several months after the effective date of the so-called "Baby Richard Amendment," the Illinois Supreme Court did not reach the constitutionality of the amendment itself, but the court did hold that application of the amendment to the Baby Richard case would violate separation of powers because the legislature would be in effect reversing the court's previous decision. *See In re* Kirchner, 649 N.E.2d 324, 337–38 (Ill. 1995).

[164] *See, e.g., In re* C.C.R.S., 892 P.2d at 260 (Lohr, J., dissenting) (noting that because trial courts applying best interests standards are reluctant to remove a child from a home after the development of psychological bonds with the family, the application of the best interests standard allows the circumvention of adoption law by parties that do not have a legally cognizable relationship with the child).

[165] *See, e.g., In re* Michael B., 604 N.E.2d 122, 131–32 (N.Y. 1992) (reasoning that absent a termination of parental rights, an award of custody to a non-biological parent would leave the child in legal and emotional limbo because the child could not be adopted).

existing adoption law, but the clash between interests embodied in these jurisdictional statutes and interests embodied in substantive adoption law should provide an opportunity for deeper reflection. "A custody dispute is more than a jurisdictional chess game in which winning depends on compliance with predetermined rules of play. A child is not a pawn."[166]

[166] E.E.B. v. D.A., 446 A.2d at 879.

BOOK REVIEW

LET US PRAY: A PLEA FOR PRAYER IN OUR SCHOOLS. By *William J. Murray*. New York: William Morrow and Company, 1995. Pp. 202, acknowledgments, introduction, bibliography. $20.00 cloth.

> Congress shall make no law respecting an establishment of religion; or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.
>
> —*U.S. Const.* amend. I

With Congress now in conservative hands, the issue of legislating prayer in public schools has returned to the political foreground. In *Let Us Pray: A Plea for Prayer in Our Schools*, William J. Murray—the plaintiff in *Murray v. Curlett*, the 1963 U.S. Supreme Court decision prohibiting mandatory school prayer[1]—decries the "secularization" of contemporary American society and the failure of our public schools to teach moral values to our children. At the heart of this problem, according to Murray, is judicial and bureaucratic misinterpretation of the U.S. Constitution's "Establishment Clause," which provides the legal framework for religious freedom in America.[2] Thus, as a step toward reinstituting civic morality, Murray proposes a constitutional amendment that would expressly permit most student-led prayer during school. Unfortunately, Murray's proposal is not only inherently coercive but also hardly essential to the task of conveying moral values to America's youth.

Murray begins his attack on secularism by recounting the fight that brought him to the forefront of the school prayer debate three decades ago, when his mother, Madalyn Murray O'Hair, sued a Baltimore school district to prevent mandatory Bible recitation in its classrooms. Murray portrays his mother as an unstable and overzealous woman driven to attack school prayer by a deep-seated hatred for God, which in turn was rooted in the refusal of Murray's Catholic father to marry her (p. 5). Moreover, Murray recalls the spontaneous nature of his mother's decision to bring the lawsuit and the negative attention he received from his classmates on its account (p. 19).

---

[1] Murray v. Curlett, 374 U.S. 203 (1963).
[2] U.S. CONST. amend. I.

Though Murray initially supported the Warren Court's decision in his case, he now views it as flawed on a variety of grounds. First and foremost, Murray attacks the Court's assumption that a secular education is a neutral one. Instead, according to Murray, secularism imposes its own morally relativistic value system upon schoolchildren. To establish that secularism amounts to a distinct moral code, Murray cites recent academic criticism of the notion of value-free neutrality and, more fundamentally, the works of famous thinkers such as Albert Einstein and Sigmund Freud. Specifically, Murray claims that Einstein's theory of relativity and Freud's theories regarding the subconscious undermine the concept of value-free neutrality by demonstrating the inability of individuals to disregard their own assumptions and values (pp. 41–42). In requiring that state action reflect secular purposes to pass Establishment Clause muster, the Court thus failed to recognize that secularism is a value system in itself, indeed one that is "overwhelmingly hostile to religion" (p. 49).

Moreover, Murray attacks the Court's interpretation of the historical and intellectual foundations of religious freedom in America. In doing so, he describes how, during the sixteenth and seventeenth centuries, the English government oppressed nonconformist religious groups and how these groups established similarly intolerant governments upon emigrating to the New World (pp. 55–74). Murray views such blatant intolerance as precisely the evil that the Constitution seeks to avoid. Thus, according to Murray, the Court's insistence upon secularism was "wholly unsupportable by the historical record" (p. 185).

Murray proceeds to emphasize the influence of the writings of English philosopher John Locke on those individuals who drafted our Constitution, particularly Thomas Jefferson. Murray maintains that while Locke called for the separation of church and state, he distrusted those who did not believe in God because they would accordingly lack the moral foundations essential to the preservation of society (p. 80). Most significantly, Murray cites some of Locke's writings in support of the proposition that Locke advocated religious training in schools (pp. 88–89).

Similarly, Murray contends that while Jefferson and Madison supported religious freedom, they did not desire the secularization of educational institutions. Regarding Jefferson's oft-cited "'wall of separation between church and State'" remark in a letter to the Danbury Baptist Association, Murray claims that

Jefferson intended it to apply only to Congress and not to public schools (pp. 89–90). Murray also suggests that Jefferson believed that strict separation of church and state was appropriate only when it promoted religious freedom "and not otherwise" (p. 93). Furthermore, as he did in the case of Locke, Murray characterizes Jefferson's statements that an important purpose of schools was to educate citizens with a sense of morality and that "'true religion is morality'" as together indicative of a belief on his part that religion belongs in public schools (p. 96).

Having pointed out what he sees as the fundamental flaws with the Court's decision in his case, Murray proceeds to trace how the same flaws have plagued modern Establishment Clause jurisprudence in general. For example, Murray criticizes the 1947 *Everson v. Board of Education* decision, in which the Court held that the Establishment Clause prohibited the government from aiding even all religions simultaneously,[3] as fundamentally at odds with both the Founders' intent and contemporary practices such as prayer in Congress and the payment of the salaries of congressional chaplains with tax revenues (p. 117). Likewise, Murray finds that even the Rehnquist Court, or at least a majority thereof, sometimes has naively equated secularism with neutrality. To illustrate this point, he mentions the 1992 *Lee v. Weisman* opinion, in which the Court held unconstitutional a non-denominational benediction offered at a graduation ceremony because it failed to reflect a secular purpose (p. 127).[4] In fact, Murray sees the Court's decisions from *Everson* to *Weisman* as stepping beyond shortsighted endorsement of secularism as essential to neutrality to evincing an outright hostile attitude toward religious activity (p. 128).

Clearly disturbed by these rulings and policies, Murray raises the specter of censorship and coercion, suggesting that suppression of religion in public schools violates the constitutional guarantee of free speech, ironically enumerated in the same amendment that contains the Establishment Clause (p. 128).[5] In a related vein, he complains that coerced secularism has replaced coerced prayer, both of which he sees as "distasteful—and unconstitutional" (p. 130). More concretely, Murray states that

---

[3] Everson v. Bd. of Educ. of Ewing, 330 U.S. 1 (1947).
[4] Lee v. Weisman, 505 U.S. 577 (1992).
[5] U.S. CONST. amend. I.

censorship and coercion are unacceptable whether imposed by "religious zealots" or "narrow-minded secularists" (p. 132).

Murray bemoans not only the rationales offered for solely secular education but also, and especially, what he believes to be the effects of such a system. Murray sees secular education as having imposed a relativistic, morally void framework on society. He accordingly claims that, in the guise of neutrality, government-mandated moral relativism has actually undermined parental moral instruction and ultimately enshrined an ethic of self-gratification (p. 167). As evidence of this effect, Murray points to increases in juvenile alcohol and drug abuse, teen pregnancy, and violent crime since the removal of prayer from public schools in the early 1960s (pp. 159–60).

Murray thus calls for the reinstatement to educational curricula of values he views as inspired by religion, specifically honesty, integrity, loyalty, fairness, altruism, justice, and compassion, as a means of avoiding the previously mentioned negative consequences of "secular" relativism (p. 166). For Murray, returning prayer to the schools would be one important step in the right direction, though he deems it "overly simplistic" to expect that doing so will in itself solve "our national disease" (p. 163). In support of his proposition that teaching moral values through public education would yield positive results, Murray refers to studies indicating that test scores from religious schools are higher than those from public schools (p. 177).

Murray enumerates three principles that he believes should guide a comprehensive effort to inculcate morality in our public schools. First, he calls for the acknowledgment and teaching of values supposedly common to all Americans. Second, Murray would have the judiciary and legislatures return control over school curricula to local communities. Third, he maintains that religion should be allowed into public schools to compete with secularism and relativism on at least an "equal footing" (p. 175). In accordance with these principles, Murray closes *Let Us Pray* by proposing that the Constitution be amended to allow prayer in public schools.

By reference to six points, Murray delineates the ideal structure of such an amendment. First, the amendment should not mandate school prayer but instead permit it only when initiated by students and accommodated by local school districts. Second, the amendment should limit the power of the federal government to interfere with local decisions to permit voluntary prayer. Third,

and seemingly repetitive of the first point, the amendment should not mandate school prayer. Fourth, the amendment should emphasize that the government does not necessarily endorse religion even though it permits student-led religious activity in state-supported educational institutions. Fifth, the amendment should allow more than simply "a moment of silence," since, in Murray's opinion, the Constitution "already protects our right to engage in silent prayer anywhere" (p. 189). Finally, the amendment should seek to restore lost rights, not to confer new ones (pp. 188–90).

In defending his proposal, Murray flatly denies that it is inherently coercive (p. 193). While admitting that some students might "feel discomfort" in a classroom engaged in prayer, Murray asserts that such discomfort is hardly more important than "matters of free speech or competing values," and that exposing differences among various belief systems will enrich students' educational experiences and encourage toleration (p. 194). Finally, Murray emphasizes that his amendment would not mandate school prayer—it would merely allow local communities to permit voluntary school prayer if they so desired. Since school policies would thus reflect "the religious or non-religious character of their communities," the risk of coercion would be minimal (p. 197).

Parts of Murray's analysis do strike a sympathetic chord. His anecdotes reveal that American institutions are, at least on certain occasions, excessively cautious regarding religious activity within their purview. For example, Murray writes of a child prohibited from singing "Jesus Loves Me" during "sharing time" at school and of a motorist prevented from having her custom license plate read "PRAY," both by civic institutions fearing entanglement with religion (p. 135).

Citing a few instances of excessive cautiousness, however, does not make for solid argument in support of prayer in public schools. Murray's suggestion that the task of instilling positive values in students requires school prayer, a belief that in turn fuels his assertion that both Locke's and Jefferson's emphasis on moral instruction indicates their support for such a policy, is highly problematic. While religious instruction does highlight the "non-controversial" values that Murray deems essential to societal stability and productivity, a well-designed secular curriculum will likely accomplish similar results. Murray's fundamental mistake is to equate extreme relativism with secular edu-

cation—while the former may occur under the auspices of the latter, it need not do so.

Indeed, although virtually all religions cherish the basic values mentioned by Murray, differences between and within faiths persist on other fronts, particularly with respect to issues involving reproductive rights, the role of women in society, and the morality of homosexuality. Murray does nothing to assure the reader that allowing the sacred back into public schools will not ultimately result in the endorsement of one religion's 'controversial' values over those of another. Accordingly, well-designed secular education seems a better mechanism for the conveyance of basic moral values than does religious instruction.

Furthermore, despite Murray's claims to the contrary, coercion under his proposed constitutional amendment would be both substantial and more marginalizing than that associated with secular education. By instilling non-controversial moral values yet leaving it to parents and the clergy to teach more contentious religious or cultural beliefs, well-designed secular education protects religious minorities against coercion without subjecting devout majorities to a hostile value system. On the other hand, Murray offers little support for his flat assertion that his school prayer proposal is not inherently coercive and even admits that one of its likely effects would be to discomfort some children. Local control hardly solves this problem, for even the smallest minorities in otherwise homogeneous communities deserve freedom from religious coercion in their public schools. Thus, on balance, secular education's limits on the free speech rights of religious students seem a fair price to pay in order to avoid more marginalizing coercion of members of minority faiths.

To Murray, permitting prayer in public schools would serve to reintroduce a lost morality and thereby forestall drug addiction, teen pregnancy and crime, all among the scourges of modern American society. Murray fails, however, to establish that such an approach is either consistent with the intent of the drafters of the Establishment Clause, necessary to remedy America's pressing social problems, or non-coercive. In reality, school prayer under virtually any framework would prove highly marginalizing for many public school students. Well-designed secular educational curricula thus seem a better path toward reinstituting civic morality.

—Alex Potente