HARVARD JOURNAL
on
LEGISLATION

VOLUME 28, NUMBER 1 WINTER 1991

ARTICLES

STATUTE

NOTE

COMMENT

RECENT DEVELOPMENTS AND LEGISLATION

HARVARD JOURNAL ON LEGISLATION
Harvard Law School
Cambridge, MA 02138

Phone: (617) 495-4400

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 decisionmaking. The JOURNAL is especially interested in publishing articles that examine a public policy problem of nationwide significance and propose legislation to resolve it.

The JOURNAL welcomes unsolicited manuscripts. Authors interested in being published in the JOURNAL should send their manuscripts and all related correspondence to: Coordinating 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 1991 volume (Volume 28) is $24.00 domestic rate; $28.00 foreign surface rate; $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 28 will be honored until January 1992.

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 placing 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*.

ARTICLE

LIMITATIONS OF ACTION UNDER THE FTCA: A SYNTHESIS AND PROPOSAL

KENT SINCLAIR*
CHARLES A. SZYPSZAK**

The Federal Tort Claims Act creates a limited waiver of the federal government's sovereign immunity, by allowing parties to sue government employees for torts committed in the course of their official duties. The Act's statute of limitations currently requires that claims be brought within two years of accrual. However, neither the Act nor the legislative history defines or offers guidelines for determining when tort claims accrue and, correspondingly, when the limitation period commences.

The authors argue that the present statute of limitations reflects a simple conception of government torts, in which the victim immediately becomes aware of his injury and its cause. Consequently, it may unfairly bar plaintiffs where the existence or origin of an injury cannot reasonably be discovered until after the limitation period has expired. The authors consider a due diligence standard of accrual, formulated by the Supreme Court in United States v. Kubrick, *which partially relaxes the strict time bar in medical malpractice actions. They survey subsequent judicial treatment of the* Kubrick *standard in various contexts for which determining the point of claim accrual is complex. Finally, the authors propose an amendment of the statute of limitations that embodies the* Kubrick *standard for all such situations.*

The 1946 Federal Tort Claims Act ("FTCA" or "the Act") creates a limited waiver of the United States' sovereign immunity from liability arising out of the tortious conduct of its employees.[1] Under section 2401(b) of the Act, a tort victim has two years to bring a suit after the cause of action accrues.[2] By restricting the time within which tort claims can be brought, this

---

* Professor of Law and Associate Dean, University of Virginia. J.D., 1971, University of California at Berkeley, Boalt Hall. The authors wish to thank Timothy Webster and Anne Gaines, University of Virginia School of Law, Class of 1991 and 1992, respectively, for research assistance. Professor Sinclair expresses gratitude for support from the Herbert Research Fellowship.

** Associate at Orr and Reno, Concord, N.H. J.D., 1986, University of Virginia. Instructor, New Hampshire Technical Institute.

[1] Act of Aug. 2, 1946, ch. 753, tit. IV, 60 Stat. 842 (codified as amended at 28 U.S.C. §§ 1346(b), 1402, 2401(b), 2402, 2411, 2412, 2671–2680 (1988)).

[2] 28 U.S.C. § 2401(b) (1982) provides:

> A tort claim against the United States shall be forever barred unless it is presented in writing to the appropriate federal agency within two years after such claim accrues or unless action is begun within six months after the date of mailing, by certified or registered mail, of notice of final denial of the claim by the agency to which it was presented.

provision encourages prompt claims and, in effect, restricts subject matter jurisdiction.

No general theory has rationalized the principal intellectual problem inherent in the FTCA's limitation doctrine: determining the point at which the cause of action accrues and the statutory period begins to run. In part the absence of theory stems from the negligence model that motivated the passage of the Act in the first place, a model typified by the image of a postal truck crashing into a plaintiff's vehicle.[3] Such fact scenarios leave no doubt that a tort has been committed. In addition, they raise no conceptual difficulties as to when the tort was committed and by whom, whether damage was caused thereby, and whether the person operating the defendant's vehicle was a government employee.

Although the legislative history of the Act strongly suggests that the Act was designed to address such "garden-variety tort suits,"[4] the explicit language of the FTCA reaches further and embraces any tort actionable under state law in the jurisdiction where the conduct occurs. As a result, the FTCA applies to a wide variety of factual circumstances, comprising more complex, attenuated, and unperceived conduct. In such situations, the accrual dates for the torts involved are difficult to determine, and the operation of the statute's two-year limitation period is therefore problematic.

Three important categories of circumstances embody the doctrinal issues that arise in determining the accrual dates for computing limitation periods; each may have considerable practical impact. In the first category, the tort is undiscoverable for a period of time, due to either the conduct of the defendant or simply the nature of the tortious event. In the second category, one or more tortfeasors actively conceal the tort from the putative plaintiff. In the third category of circumstances, the tort is an ongoing tort for which the putative plaintiff defers bringing suit. The following examples reflect the breadth and significance of the problem of determining the appropriate limitation period.

1. *Late Discovery*

(a) The plaintiff is the victim of a tort arising from past exposure to a toxic substance.[5]

---

[3] *See infra* note 28 and accompanying text.
[4] C.P. Chemical Co. v. United States, 810 F.2d 34, 36 (2d Cir. 1987).
[5] *See, e.g.*, Schnurman v. United States, 490 F. Supp. 429 (E.D. Va. 1980) (exposure to mustard gas during U.S. Navy experiment).

(b) The plaintiff is in a coma for more than two years after government physicians committed malpractice in treatment.[6]

2. *Concealment*[7]

(a) The plaintiff is the victim of a secret prison medical test that results in physical harm.[8]
(b) The plaintiff is the unknowing subject of an improper investigation by law enforcement personnel or other government agents, carried out using electronic surveillance, undercover agents, informants, or other techniques calculated not to come to the investigatee's attention.[9]
(c) The plaintiff is injured during a medical test negligently conducted by government physicians who purposely suppress the plaintiff's medical records to prevent discovery of the injury and its cause.[10]

3. *Continuous Course of Conduct*

The plaintiff owns farmland that is subject to annual flooding as a result of the construction and operation of a military base located on an adjacent parcel of property.[11]

At the heart of the current unsatisfactory understanding of the FTCA limitation period's operation lies the Supreme Court's decision in *United States v. Kubrick*,[12] which held that a plaintiff's medical malpractice action accrued when he became aware of both the existence and cause of his injury. Many courts have

---

[6] Clifford v. United States, 738 F.2d 977 (8th Cir. 1984) (statute did not run on comatose patient; period began to run only when guardian was appointed for patient, and when guardian had requisite knowledge of patient's injury and a legal duty to act for the patient). *See also* Washington v. United States, 769 F.2d 1436 (9th Cir. 1985) (cause of action accrued when patient died, not earlier when patient went into coma; suit held timely when instituted by survivors promptly after the death of the victim); Dundon v. United States, 559 F. Supp. 469 (E.D.N.Y. 1983) (although the date of accrual began before the decedent entered a coma, the statute of limitations was tolled during the period that the decedent lay comatose, since his mental condition, allegedly caused by his physicians, directly prevented him from understanding the nature and cause of his injuries).

[7] *See infra* notes 187–213 and 300–301.

[8] *See* Cain v. United States, 643 F. Supp. 175 (S.D.N.Y. 1986); Scott v. Casey, 562 F. Supp. 475 (N.D. Ga. 1983).

[9] *See, e.g.*, Hobson v. Wilson, 737 F.2d 1 (D.C. Cir. 1984), *cert. denied sub nom.* Brennan v. Hobson, 470 U.S. 1084 (1985).

[10] *See* Harrison v. United States, 708 F.2d 1023 (5th Cir. 1983).

[11] *See* Korgel v. United States, 619 F.2d 16 (8th Cir. 1980).

[12] 444 U.S. 111 (1979).

deemed *Kubrick* to be limited to medical malpractice actions.[13] Some courts, however, have extended its approach to other forms of tortious conduct.[14] We propose that *Kubrick*'s approach offers a sound analytical basis for analyzing all circumstances arising under the Act, therefore ensuring consistency in the circuits.[15] Although the decision is widely misread by lower courts that ignore its factual context,[16] *Kubrick* contains the seeds of a general theory of the accrual of causes of action that can provide the basis for a coherent policy applicable to all factual contexts in which the Act is applied.

In Part I of this Article, we briefly describe the prevailing standards used by courts in applying the limitation period prescribed by the FTCA. In Part II we examine the way these standards are used in various types of cases. In Part III we propose that a simple discovery and reasonable diligence standard, such as that employed in medical malpractice cases under the Act,[17] should govern in all of the troublesome situations illustrated by the above hypotheticals. Furthermore, the coherence of such a unified rule strongly urges its use in a variety of other situations where strictly applying the FTCA's limitation period currently produces harsh results. In Part IV we offer an

---

[13] Cases purporting to limit the *Kubrick* rationale to medical malpractice cases include Gross v. United States, 676 F.2d 295, 300 (8th Cir. 1982) (in a claim for intentional infliction of emotional distress arising out of a fraudulent attempt to deny the plaintiff participation in a feed grain program, the court held that "*Kubrick* involved a medical malpractice action, not a continuing tort. Where the tortious conduct is of a continuing nature, the *Kubrick* rule does not apply . . . ."), and Ware v. United States, 626 F.2d 1278, 1284 n.4 (5th Cir. 1980) (rejecting application of "the medical malpractice accrual test articulated in *United States v. Kubrick*" to an FTCA action in which the government misdiagnosed the plaintiff's cattle as tubercular and consequently destroyed them).

See also Herrera-Diaz v. United States, Dep't of Navy, 845 F.2d 1534, 1536 (9th Cir. 1988) ("Tort claims usually accrue at the time of a plaintiff's injury. In medical malpractice actions under the FTCA, however, a claim does not accrue until a plaintiff discovers both the injury and its cause . . . ."), *cert. denied*, 488 U.S. 924 (1988).

[14] Indeed, *Kubrick* has been taken by some as projecting a "federal" approach to accrual of causes of action, applicable in other contexts. *See, e.g.*, Sowers v. Bradford Area School Dist., 694 F. Supp. 125, 136 (W.D. Pa. 1988) (applying *Kubrick* discovery rule to determine accrual of 42 U.S.C. § 1983 action, citing cases), *aff'd*, 869 F.2d 591 (1989), *cert. denied sub nom.* Smith v. Sowers, 110 S. Ct. 840 (1990); *see also* Urie v. Thompson, 337 U.S. 163 (1949) (applying similar rule under the Federal Employers's Liability Act); *see generally* S. NAHMOD, CIVIL RIGHTS AND CIVIL LIBERTIES LITIGATION § 4.15 at 255 (2d ed. 1986) (construing *Kubrick* to indicate the approach for civil rights suits under 42 U.S.C. § 1983).

[15] For example, the adoption of such an approach will resolve differences among the circuits on the issue of whether the diligence discovery approach applies to wrongful death cases under the FTCA. See *In re* Swine Flu Prods. Liab. Litig. (Sanborn v. United States), 764 F.2d 637 (9th Cir. 1985) (discussing divergence of view among the circuits on this issue and citing cases).

[16] *See infra* text accompanying notes 98–104.

[17] *See infra* notes 87–104.

amendment to the FTCA that would effect the doctrinal unification proposed here.

I. The FTCA Limitation Provision: History and Interpretations

In a time when the federal government is nearly omnipresent in American society, the Federal Tort Claims Act is a critical remedial provision defining the substantive rights of citizens to bring suit against the government for its tortious conduct. The availability of an FTCA remedy, however, often turns on the operation of the statute's limitation period, a feature that is sometimes overlooked or dismissed as a pro forma procedural hurdle. While appearing straightforward, the FTCA's limitation provision is surrounded by confusion and conflict, as demonstrated by examining its history and the available guidelines for its interpretation.

A. *Legislative History*

The Federal Tort Claims Act was enacted in 1946 as part of a legislative package designed to reorganize and reallocate congressional responsibilities,[18] including its responsibility for the tortious conduct of government agents and employees. Up to that point, the federal government had not waived its sovereign immunity against tort liability, although it had long before waived immunity from suit for other types of legal actions.[19] In an effort to end the burdensome and anachronistic procedure of securing relief from torts committed by the federal government through private bills,[20] Congress vested the United States District Courts with article I jurisdiction[21] to hear specified tort

---

[18] Legislative Reorganization Act of 1946, ch. 753, 60 Stat. 812.

[19] *See, e.g.*, Act of Mar. 3, 1887, ch. 359, 24 Stat. 505 (waiver of sovereign immunity for claims based upon the Constitution, acts of Congress, any regulation of an executive department, or express or implied contracts with the government of the United States). *See generally* Holtzoff, *The Handling of Tort Claims Against the Federal Government*, 9 L. & CONTEMP. PROBS. 311 (1942) (discussing congressional waivers of sovereign immunity prior to the FTCA).

[20] By the 1940's, Congress was receiving approximately 2000 private claims bills each year. *See* S. REP. No. 1400, 79th Cong., 2d Sess. 30–31 (1946). These bills were processed through the Claims Committees of both houses with an eventual success rate of roughly 25%. *Id.* Not all of these claims arose out of tortious conduct; however, a report on an earlier draft of the FTCA estimated that up to 60% of all claims would be handled through this Act. *See* H.R. REP. No. 2800, 71st Cong., 3d Sess. 2 (1931).

[21] U.S. CONST. art. I, § 8.

claims[22] and concurrently dissolved its own Claims Committees.[23] Partial relief from the private claims burden, as well as the creation of a right to recovery from government torts, appear to be the overarching purposes of the legislation.[24]

Although the FTCA was the product of twenty years of legislative proposals, hearings, reports, enactments, and vetoes,[25] the statute of limitations provision of the Act[26] was rarely at issue. The few discussions that did concern the provision shed little light on the intended meaning of "accrual" for the purposes of the Act.[27] Lack of concern over this issue may have stemmed from the fact that Congress envisioned the Act as addressing mundane, common law torts, most of which arise out of specific, straightforward incidents such as automobile accidents.[28] Furthermore, although the Act has been revised several times, the amendments and their accompanying legislative history do not clarify the lawmakers' intentions. This obscurity of intention is further clouded by the legislative history of the final version of the Act, which differs from that of the earlier versions and, therefore, casts doubt on the prior history.[29] To a great extent, therefore, interpretation of this section of the Act has been left to the judiciary.

---

[22] 60 Stat. at 843–44.

[23] 60 Stat. at 818, 826–27. All remaining domestic claims responsibilities were transferred to the Judiciary Committees of each house. *Id.* Further, Congress banned the consideration of private bills for which a claim could be pursued under the FTCA. *Id.* at 831. Despite this ban, claims which fall within an exception to the FTCA can still be presented to Congress. *See* HOUSE COMM. OF THE JUDICIARY, 101ST CONG., 1ST SESS., SUPPLEMENTAL RULES OF PROCEDURE FOR PRIVATE CLAIMS BILLS 2–3 (Comm. Print 1989) (First Amendment to the Constitution guarantees the right "to petition the Government for a redress of grievances"); *see also infra* note 257 and accompanying text.

[24] *See Tort Claims Against the United States: Hearings on H.R. 7236 Before Subcomm. No. 1 of the House Comm. on the Judiciary*, 76th Cong., 3d Sess. (1940) [hereinafter *Hearings on H.R. 7236*]; *see also* H.R. REP. No. 2800, *supra* note 20, at 3; Armstrong & Cockrill, *The Federal Tort Claims Bill*, 9 LAW & CONTEMP. PROBS. 327 (1942).

[25] *See* SEN. REP. No. 1400, 79th Cong., 2d Sess. 30–31 (1946); *see also* Comment, *Federal Tort Claims Act*, 56 YALE L.J. 534 (1946–47).

[26] 60 Stat. 845, § 420.

[27] *See Kubrick*, 444 U.S. at 119.

[28] *See, e.g.*, 89 CONG. REC. A4185 (Extension of Remarks of Hon. Paul Stewart) (describing the need for a remedy for "the ordinary common-law tort" committed by a government employee, such as the victim of a collision with a postal truck); H.R. REP. No. 2245, 77th Cong., 2d Sess. 7 (1942) (noting the particular importance of giving a right to sue "in respect to such torts as negligence in the operation of vehicles"); *see also* C.P. Chemical Co. v. United States, 810 F.2d at 36–37 (2d Cir. 1977) (observing that the focus of the FTCA was on liability for "negligence in the operation of vehicles").

[29] *See* United States v. Yellow Cab, 340 U.S. 543, 549–50 (1951) ("[t]he reports [of the final FTCA bill] omitted previous discussions which tended to restrict the scope of the Tort Claims bill").

The original language of the FTCA's statute of limitations appears to be based on a pre-existing provision allowing district courts original jurisdiction in suits brought against the United States.[30] The statute had disallowed any suit against the federal government unless it had been brought "within six years after the right [for which the claim had been made] accrued."[31] As enacted, the FTCA's limitation provision specified a much shorter time for bringing tort actions than had been allowed for other actions brought against the United States. It provided that claims against the United States under the FTCA would "be forever barred" unless an action was brought "within one year after such claim accrued or within one year after the date of enactment of th[e] Act, whichever [was] later."[32] What little interpretive material exists suggests that this short period was intended to encourage the prompt presentation of claims and thereby prevent injustice to the government resulting from the loss of critical evidence.[33] With minor stylistic amendments,[34] the statute of limitations provision was recodified in 1948 as 28 U.S.C. § 2401(b).[35]

In 1949, Congress lengthened the limitation period an additional year.[36] Lawmakers sensed that the prior one-year period

---

[30] Judiciary Act of 1911, ch. 231, § 24, para. 20, 36 Stat. 1091, 1093 (repealed 1946). The six-year statute of limitations for actions against the government not sounding in tort can now be found at 28 U.S.C. § 2401(a)(1988).

[31] *Id.*

[32] 60 Stat. 845, § 420 (codified at 28 U.S.C. § 942 (1946)). Also, denials of claims for less than $1,000 that had been brought pursuant to the administrative claim provision of the Act could be challenged within an additional six-month period. *Id.*

[33] *See* H.R. REP. No. 2428, 76th Cong., 3d Sess. 5 (1940). Hearings on earlier versions of the FTCA had provoked the following comments:

> I realize the handicap an agency would be under if it were sued for an accident as much as a year after it happened, without any notice. Witnesses would have gone to the four winds, and there would be no opportunity to present a proper defense.

Tort Claims Against the United States: Hearings on S. 2690 Before Subcomm. of the Sen. Comm. on the Judiciary, 76th Cong., 3d Sess. 22 (1940) (remarks of Herbert Bingham, ABA). "[One year] is necessary for the purpose of protecting the interests of the government." *Id.* at 38 (remarks of Alexander Holtzoff, Dep't of Justice). Mr. Holtzoff also noted that in cases of hardship caused by the one-year period, a claimant could still seek relief from Congress through a private bill. *Id.* at 38, 47. Indeed, he expected that this would occur and noted that it had occurred in other areas already. *Hearings on H.R. 7236*, *supra* note 24, at 21.

No language in the numerous hearings and reports specifically refers to what is meant by the term "accrues." Rather, the recorded statements address the length of the period in which to bring a claim once that claim is actionable.

[34] *See* S. REP. No. 1559, 80th Cong., 2d Sess. 10 (1948).

[35] Act of June 25, 1948, ch. 646, 62 Stat. 971.

[36] Act of Apr. 25, 1949, ch. 92, 63 Stat. 62. The only legislative history accompanying this Act is H.R. REP. No. 276, 81st Cong., 1st Sess., *reprinted in* 1949 U.S. CODE

was "too short and tend[ed] toward injustice in many instances,"[37] and that it should correspond more closely with the period provided in most state statutes of limitations for torts and in other federal statutes of limitations.[38] Specifically, Congress explained that the one-year period was "unfair to some claimants who suffered injuries which did not fully develop until after the expiration of the period for making a claim."[39] Although Congress recognized that the extension of the statutory period to two years would expand opportunities for bringing suit under the FTCA,[40] it did not intend the amendment to be viewed as a congressional effort to "encourage delay in the enforcement of a claimant's rights or to harass the Federal agencies in the defense against such suits."[41] The amendment merely extended by one year the time to file a claim after that claim had already accrued. The amendment, however, did not address the meaning of accrual.[42]

Congress has passed a variety of other amendments to section 2401 and its predecessors since 1911, but none bears on the

CONG. & ADMIN. NEWS 1226–29. A parallel bill had been introduced in the Senate and had been reported favorably. *See* S. REP. No. 135, 81st Cong., 1st Sess. (1949). A similar bill had been introduced in the previous session of Congress and had evoked a similar report. *See* H.R. REP. No. 1754, 80th Cong., 2d Sess. (1948). As originally introduced, the bill called for the expansion of the statute of limitations to three years. H.R. 4682, 80th Cong., 2d Sess. (1948).

[37] H.R. REP. No. 276, *supra* note 36, at 2.

[38] *Id.* at 1228–29. The committee found that the average state limitation period for torts was 2.92 years, and the average period for other federal causes of action, for instance certain admiralty actions, was approximately 2.2 years. *Id.*

[39] *Id.* This statement appears to imply that a claim accrues when an injury occurs rather than when it is fully developed; otherwise there would be no reason for extending the limitation period. *See* Brief for the United States at 25, United States v. Kubrick, 444 U.S. 111 (1979) (No. 78-1014). However, the statement could mean that once a claim has accrued, the limitation period is not tolled by the fact that the resulting injury is not fully manifested. Brief for Respondent at 18–19 n.12, United States v. Kubrick, 444 U.S. 111 (1979) (No. 78-1014). Yet this language leaves ambiguous exactly at what stage of an injury, prior to "complete development," a claim may accrue for purposes of the Act.

[40] *See* H.R. REP. No. 276, *supra* note 36, at 3.

[41] *Id.* at 4.

[42] A further amendment in 1966 gave § 2401(b) its current form by altering the FTCA's administrative claim requirements. Act of July 19, 1966, Pub. L. 89-506, § 7, 80 Stat. 307. In Brief for the United States, *supra* note 39, at 26, the government argues that two other bills passed on the same day as the 1966 amendment elucidate Congress's understanding of how causes of action accrue for statute of limitations purposes. *See* 28 U.S.C. §§ 2415, 2416 (1988). These two statutes establish statutes of limitations for actions brought by the United States and adopt a diligence discovery rule. *Id.* The government argued that if Congress intended a similar rule under the FTCA, it would have provided as such. *Id.* at 27. This ignores the fact, however, that the 1966 amendments to the FTCA were focused on the administrative claim system, and the legislative history demonstrates that the statute of limitations issue was not addressed. *See generally* 1966 U.S. CODE CONG. & ADMIN. NEWS 2515.

definition of the period currently applied to such claims. Most significantly for purposes of this analysis, none of the other legislative materials informs our understanding of when a claim is deemed to accrue for limitation purposes or what circumstances, if any, extend or "toll" the running of the limitation period.[43]

Of course, Congress need not and cannot define every legislative term. Those having unambiguous usage should speak for themselves. But "accrual," at least in the statute of limitations context, had a broad range of connotations at the time the Act was passed.[44] Consequently, courts have had to choose which meaning to apply.

B. *Guides to Construction*

The courts have strictly applied section 2401(b),[45] treating it as a typical statute of repose.[46] As the Supreme Court has noted:

---

[43] A full roster of the legislative developments in this area follows: Act of Mar. 3, 1911, ch. 231, § 24, para. 20, 36 Stat. 1087, 1093 (origins of present six-year period found in 28 U.S.C. § 2401(a) (1988)); Act of Nov. 23, 1921, ch. 136, § 1310(c), 42 Stat. 227, 311; Act of June 2, 1924, ch. 234, § 1025(c), 43 Stat. 253, 348; Act of Feb. 24, 1925, ch. 309, 43 Stat. 972; Act of Feb. 26, 1926, ch. 27, § 1122(c), 44 Stat. 9, 121; Act of Aug. 2, 1946, ch. 753, tit. IV, § 420, 60 Stat. 812, 845 (first FTCA provision, one-year period); Act of June 25, 1948, ch. 646, 62 Stat. 971; Act of Apr. 25, 1949, ch. 92, § 1, 63 Stat. 62 (providing present two-year limitation period); Act of Sept. 8, 1959, Pub. L. 86-238, § 1(3), 73 Stat. 471, 472; Act of July 18, 1966, Pub. L. 89-506, § 7, 80 Stat. 306, 307; Act of Nov. 1, 1978, Pub. L. 95-563, § 14(b), 92 Stat. 2143, 2389.

[44] "Accrue" was commonly defined to mean "to arise, to happen, to come into force or existence; to vest; as in the phrase, 'The right of action did not *accrue* within six years.'" BLACK'S LAW DICTIONARY 29 (3d. ed. 1931); *see also* BALLENTINE'S LAW DICTIONARY 16 (2d ed. 1948) ("[a]s applied to a cause of action, the word means to arrive; to commence; to come into existence; to become a present and enforceable demand").

Restatement of Torts explains that "the statute does not usually begin to run until the tort is complete . . . . [It is] ordinarily not complete until there has been an invasion of a legally protected interest of the plaintiff." RESTATEMENT OF TORTS § 899 (1939). In regard to knowledge of the tort, *Restatement of Torts* noted: "it is still true that in many of the States that, in the absence of fraud or concealment of the cause of action, the statutory period runs from the time the tort was committed although the injured person had no knowledge or reason to know of it." However, several states had adopted a discovery-style rule for medical malpractice cases. *Id.* at comment e. Because the federal courts have determined that federal law controls the definition of accrual, however, state constructions of that term cannot be imputed to Congress.

[45] *See* Soriano v. United States, 352 U.S. 270, 276 (1957); United States v. Sherwood, 312 U.S. 584 (1941); Moll v. US Life Title Ins. Co. of New York, 654 F. Supp. 1012, 1021 (S.D.N.Y. 1987); Hammond v. United States, 388 F. Supp. 928, 930 (E.D.N.Y. 1975) (citing numerous cases).

[46] *See generally* W. FERGUSON, THE STATUTES OF LIMITATION SAVING STATUTES (1978).

> [A]lthough affording plaintiffs what the legislature deems a reasonable time to present their claims, [statutes of limitations] protect defendants and the courts from having to deal with cases in which the search for truth may be seriously impaired by the loss of evidence, whether by death or disappearance of witnesses, fading memories, disappearance of documents, or otherwise.[47]

Courts have strictly construed section 2401(b) and have largely refused to create exceptions to its operation.[48] Some courts have attributed their reluctance to bend the temporal bar to their belief that they lack power to create equitable waivers of its application.[49]

In construing the FTCA's limitation provision, the Supreme Court has said that courts should not "extend the waiver beyond that which Congress intended."[50] This strict approach is consistent with the Supreme Court's assessment of what the Act's legislative history commands.[51] In accord with that approach, federal courts have consistently refused to make the usual equitable exceptions to the operation of the statute of limitations for cases brought under the FTCA. For instance, courts have held that neither minority status,[52] mental incapacitation,[53] nor state of war will toll the limitation period for a claimant.[54] Thus courts have generally refused to read section 2401(b) provisions

---

[47] *Kubrick*, 444 U.S. at 117. In *Kubrick*, the plaintiff brought a tort claim against the government alleging mistreatment at a Veterans Administration hospital. The Supreme Court held that the plaintiff's claim accrued when he knew of both the existence and cause of his injury, but that he need not have been aware of negligence, for he was on notice to seek expert advice once the injury and its cause were discovered. *See infra* text accompanying notes 87–97 for a full discussion of *Kubrick*.

[48] *See* Block v. North Dakota Bd. of Univ. and School Lands, 461 U.S. 273 (1983), where the Supreme Court held that a state must comply with conditions on waivers of sovereign immunity when bringing a suit against federal officials. The Court said: "[W]hen Congress attaches conditions [such as a statute of limitations] to legislation waiving the sovereign immunity of the United States, those conditions must be strictly observed, and exceptions thereto are not to be lightly implied." *Id.* at 287.

[49] *See* Caidin v. United States, 564 F.2d 284, 286 (9th Cir. 1977); *see also* Camire v. United States, 489 F. Supp. 998 (N.D.N.Y. 1980); Sangeminio v. Zuckerberg, 454 F. Supp. 206 (E.D.N.Y. 1978). *But see infra* text accompanying notes 128–138 (discussing consideration by federal courts of equitable concerns in applying the FTCA's statute of limitations).

[50] *Kubrick*, 444 U.S. at 117–18. *See generally Block*, 461 U.S. at 287; Lehman v. Nakshian, 453 U.S. 156, 161 (1981).

[51] *See, e.g.*, Kosak v. United States, 465 U.S. 848, 853–55 (1984) (courts construing subsections of the FTCA should identify circumstances only within the words and reasons of the provisions involved).

[52] Hoch v. Carter, 242 F. Supp. 863, 865 (S.D.N.Y. 1965).

[53] Barren v. United States, 839 F.2d 987 (3d Cir. 1988), *cert. denied*, 109 S.Ct. 79 (1988).

[54] Soriano v. United States, 352 U.S. 270, 275–76 (1957).

broadly, even when a narrow construction has caused harsh results[55] and risked a denial of the relief Congress intended to provide through the immunity waiver.[56]

C. *Law Governing Accrual Determination*

State judicial construction of similar state law provisions provides an obvious and potentially fertile source of guidance for interpreting the FTCA's limitation provision. Indeed, issues arose early in the judicial interpretation of section 2401(b) over whether state or federal law controlled the determination of when a claim accrued under the Act. The FTCA is uniquely structured to draw the substantive tort causes of action from state law, by providing that "the district courts . . . have exclusive jurisdiction of civil actions on claims against the United States . . . where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred."[57]

This structure seems to suggest that where tort liability in a particular state is circumscribed by that state's rules for accrual of causes of action, those rules should also be used in applying the FTCA. Such an interpretation would not abrogate section

[55] *See* Bailey v. United States, 642 F.2d 344, 346–47 (9th Cir. 1981) (suit by heirs of decedent killed on gunnery range by alleged negligence of the Air Force dismissed as time-barred where claim had been mailed but not properly presented); Wollman v. Gross, 637 F.2d 544, 549 (8th Cir. 1980) (suit for alleged negligence of government employee in automobile accident time-barred despite the plaintiff's claim that he was "blamelessly ignorant" of the legal significance of an accident with a government employee); Snodgrass v. United States, 567 F. Supp. 33, 34–36 (E.D.N.Y. 1983) (claim for alleged unlawful arrest, imprisonment, and detention by federal agents time-barred despite allegations that the plaintiff had not known the agents were involved); *In re* "Agent Orange" Prod. Liab. Litig., 506 F. Supp. 757, 760 (E.D.N.Y. 1980) (in litigation against the Navy for alleged injuries from exposure to the chemical known as "agent orange," time constraints of § 2401(b) were not subject to extension or waiver).

[56] The Court in *Kubrick* cited Indian Towing Co. v. United States, 350 U.S. 61 (1955), in which the Court had stated:

> The broad and just purpose which the statute was designed to effect was to compensate the victims of negligence in the conduct of governmental activities in circumstances like unto those in which a private person would be liable and not to leave just treatment to the caprice and legislative burden of individual private laws. Of course, when dealing with a statute subjecting the Government to liability for potentially great sums of money, this Court must not promote profligacy by careless construction.

Id. at 68–69.

[57] 28 U.S.C. § 1346(b) (1988). *See also* 28 U.S.C. § 2674 (1988) ("[t]he United States shall be liable . . . in the same manner and to the same extent as a private individual under like circumstances").

2401(b); once a claim is held to have accrued according to state law, it would be untimely if nôt pursued within two years.

In contrast to the liability provisions of the Act, however, section 2401(b)'s text provides no reason to believe that it was meant to be interpreted in accordance with state law.[58] The legislative history of section 2401(b) does not aid in resolving this point.[59]

Although several early cases held that state law controlled when a claim accrued, subsequent cases have concluded that federal law controls this determination.[60] The principal consideration guiding most federal decisions on the issue is that yielding the accrual issue to state law would preclude the uniformity sought through the statute of limitations provision.[61] In the FTCA, as several courts have noted, Congress deferred to state law for tort definitions because of the technical complexity of delineating every possible ground for liability in the Act itself;[62] however, the statute of limitations was a discrete issue that Congress could and did address uniformly according to federal law.[63] One court has further reasoned that if a state rule of accrual resulted in a shorter limitation period than a federal rule, more private claims for relief would be submitted to Congress, a situation the FTCA was meant to prevent.[64] Federal law is now uniformly held to control when a claim accrues under the Act.[65]

---

[58] See *supra* note 2 for the text of § 2401(b).

[59] *See supra* text accompanying notes 18–44.

[60] *See, e.g.*, Tyminski v. United States, 481 F.2d 257, 262–63 (3d Cir. 1973); *see generally* Annotation, *Statute of Limitation Under Federal Tort Claims Act*, 29 A.L.R. Fed. 482 § 5 (1976 & Supp. 1989). Until the 1980's, the First Circuit held the view that state law governs when a claim accrues for purposes of the FTCA. *See* Hau v. United States, 575 F.2d 1000, 1002 (1st Cir. 1978). Although that court has not expressly reversed itself, in more recent cases it has referred exclusively to federal law for the definition of accrual. *See* Nicolazzo v. United States, 786 F.2d 454, 455 (1st Cir. 1986). Lower courts within that circuit have interpreted *Nicolazzo* as holding that federal law controls. *See* Santana v. United States, 693 F. Supp. 1309, 1312 n.2 (D.P.R. 1988).

[61] *See* Quinton v. United States, 304 F.2d 234, 235–40 (5th Cir. 1962); Maryland v. United States, 165 F.2d 869, 871 (4th Cir. 1947).

[62] *See, e.g.*, *Quinton*, 304 F.2d at 236.

[63] *Id.*

[64] *Maryland*, 165 F.2d at 872.

[65] *See supra* note 60. The Supreme Court's decision in *Kubrick* implicitly removes any remaining doubts on this point. Although the Court in *Kubrick* did not expressly address the choice of law issue, the accrual rule that emerged was in large part based on the Court's opinion in Urie v. Thompson, 337 U.S. 163 (1949), a case that held that federal law controls. *Kubrick* at 120 n.7. Thus the Court appeared to assume that federal law controlled in *Kubrick*. Furthermore, the holding in *Kubrick* clearly establishes the parameters of the accrual of a cause of action under the Act and thus any state rules that conflicted with the holding would not apply. *Id.* at 124–25.

Oddly, this results in the separation of the limitation doctrine from the substantive causes of action. In other contexts the Supreme Court has held that state statutes of limitations are substantive law, requiring that a court exercising diversity jurisdiction apply the state period.[66] State rules supersede federal rules in determining when the cause of action accrues and what steps are required for its commencement.[67] Under the FTCA, however, while liability is premised on state tort law, the jurisdiction exercised rests not upon diversity but federal subject matter. Therefore, the federal Act's limitation period applies and limits the substantive causes of action based on state law.

Nonetheless, state law still influences the accrual issue in one important respect: a claim cannot accrue for statute of limitations purposes until it exists, and state law determines its existence.[68] Thus state law controls when the event is defined as an actionable tort, and federal law governs when, on or after that date, a claim accrues for purposes of section 2401(b). This mechanical point does not conflict with the current FTCA statute of limitations jurisprudence, discussed below.

II. Present Operation of the Statute

Although many cases brought under the FTCA are clearly either within or without the limitation period, in other cases courts must struggle with the question of whether a plaintiff has brought suit within the statutory time limit. For cases that do not fit the notion of a simple, obvious, physical tort that usually comes to mind when envisioning an FTCA suit, the courts have developed an approach by which they assess the plaintiff's diligence (or lack thereof) in determining whether the limitation period should be applied strictly.[69] To understand why this approach is only a partial solution to the problem, the approach and its application must first be examined.

---

[66] *See* Guaranty Trust Co. v. New York, 326 U.S. 99, 112 (1945).

[67] *See, e.g.*, Walker v. Armco Steel Corp., 446 U.S. 740, 751–52 (1980) (In a diversity suit, Fed. R. Civ. P. 3 does not govern when an action is filed for purposes of Oklahoma's statute of limitations; the state's "actual service" of process requirement is substantive and must be satisfied before the limitations period is tolled).

[68] *See* Bizer v. United States, 124 F. Supp. 949, 952 (N.D. Cal. 1954); *see also* Foote v. United States, 107 F. Supp. 270, 275 (W.D. Mich. 1952).

[69] We refer hereinafter to this approach as the "diligence discovery" or "due diligence" rule.

A. *Simple, Obvious, Physical Torts*

Section 2401(b) provides that a prospective plaintiff's tort claims are "forever barred unless . . . presented in writing to the appropriate Federal agency within two years after such claim" accrued.[70] Ordinarily, a cause of action accrues when the alleged injury occurs.[71] The alleged injury is deemed to occur when the actionable conduct is complete, rather than when its effects are felt.[72] Thus, when the actionable conduct is complete, the statutory period begins to run.[73] The premise in such situations, however, is that the conduct will immediately come to the victim's attention.[74] The Supreme Court has commented that the paradigm for application of the Act as a whole is a simple traffic accident involving a government vehicle.[75]

B. *Medical Malpractice Cases—the* Kubrick *Approach*

Following enactment of the FTCA, courts began to recognize that equating accrual of a cause of action with the date of the act or omission that caused the tortious injury could potentially lead to injustice in the medical malpractice context.[76] Often, the victim of medical malpractice may not know of an injury or its cause within a short statutory period of limitations because the consequences of malpractice may take years to develop, and the character of the injury may prevent discovery by anyone but a physician.[77] A legal rule that charges an unknowing plain-

---

[70] 28 U.S.C. § 2401(b) (1988).

[71] *Kubrick*, 444 U.S. at 120.

[72] *See generally* Shockley v. Vermont State Colleges, 793 F.2d 478, 481 (2d Cir. 1986), and cases cited therein; Chardon v. Fernandez, 454 U.S. 6, 8 (1981); Delaware State College v. Ricks, 449 U.S. 250, 258 (1980).

[73] *Kubrick*, 444 U.S. at 120.

[74] *See* C.P. Chemical Co. v. United States, 810 F.2d 34, 36 (2d Cir. 1987).

[75] Kosak v. United States, 465 U.S. 848, 855 (1984) ("[o]ne of the principal purposes of the Federal Tort Claims Act was to waive the Government's immunity from liability for injuries resulting from auto accidents").

[76] *See, e.g.*, Quinton v. United States, 304 F.2d 234 (5th Cir. 1962). In *Quinton*, the plaintiff was negligently given transfusions of RH positive rather than RH negative blood in an Air Force hospital. *Id.* at 235. The plaintiff had no reason to suspect that a mistake had been made. As a direct result of the negligent transfusions, the plaintiff delivered a stillborn baby four years later. *Id.*

[77] *See* RESTATEMENT (SECOND) OF TORTS § 899 comment e (1979), *quoted in Kubrick*, 444 U.S. at 120 n.7. Also, medical malpractice claims often arise from negligent omissions, particularly failures to give adequate medical advice, which plaintiffs have no reason to suspect because they are not themselves aware of the potential alternatives. In these situations, there is likely to be a significant time delay between the omission and an event that will bring it to light.

tiff with knowledge of a malpractice cause of action for statute of limitations purposes would be akin to the early English common law fiction that a plaintiff has an actionable case as of the date of an injury notwithstanding the plaintiff's own ignorance of that injury or its cause.[78] Although the federal courts have been willing to adopt a more equitable rule at least to a limited extent,[79] the exact parameters of such a rule are still the subject of debate.

To alleviate the injustice of barring a medical malpractice action under the FTCA before a putative plaintiff even recognizes its existence, the federal courts have developed a "diligence discovery" rule of accrual in such cases. Under this approach a claim does not accrue until "the claimant discovered, or in the exercise of reasonable diligence should have discovered, the acts constituting the alleged malpractice."[80] The courts derived this rule from the Supreme Court's earlier consideration of the accrual of a cause of action under the Federal Employers' Liability Act, where it found that a plaintiff's "blameless ignorance" concerning a cause of action under the Act could not be held against him.[81] By permitting some suits that otherwise would be barred, it actually serves Congress's intent to reduce the number of private bills introduced for the purposes of resolving tort claims against the government.[82] This rule quickly spread throughout the federal courts of appeals.[83]

---

[78] *See* RESTATEMENT OF TORTS § 899 comment e (1939). Around the time the FTCA was enacted, this was in fact the rule in many states. *See generally* Lillich, *The Malpractice Statute of Limitations in New York and Other Jurisdictions*, 47 CORNELL L.Q. 339 (1962). A diligence discovery approach now prevails in a number of states. *See infra* note 279 and accompanying text.

[79] *See, e.g., Quinton*, 304 F.2d at 240.

[80] *Quinton*, 304 F.2d at 240.

[81] Urie v. Thompson, 337 U.S. 163, 171 (1949). *Urie* involved a railroad worker's claim that he contracted silicosis as a result of his employer's negligent failure to protect him adequately against silica dust. *Id.* at 165–66. While the Court held that "each intake of dusty breath" does not constitute a separate cause of action under the Act, *id.* at 170, the plaintiff had no reason to know of his condition at an earlier date and could not have an actionable injury until the "'effects of the deleterious substance manifest themselves.'" *Id.* (citation omitted). Thus his action was not time-barred since it was filed within the limitations period after he discovered his injury. *Id.*

[82] Tort claims that are barred by the FTCA's statute of limitations presumably may be introduced as private bills in Congress. *Cf.* Maryland v. United States, 165 F.2d 869, 872 (4th Cir. 1947) (construction of the limitation period is governed by federal law because state periods that are shorter would thwart the FTCA's purpose of eliminating private bills for tort claims, since plaintiffs barred by a state limitation period would still be able to submit private bills).

[83] *See generally* Hungerford v. United States, 307 F.2d 99 (9th Cir. 1962); Tyminski v. United States, 481 F.2d 257 (3d Cir. 1973); Toal v. United States, 438 F.2d 222, 224–25 (2d Cir. 1971).

By the 1970's, however, an important ambiguity in the diligence discovery rule became apparent: does the limitations period begin to run when the plaintiff knows, or should know, of an injury and its cause, or not until the plaintiff also knows that the injury was negligently inflicted? Each position has merit. On the one hand, a cause of action should not accrue until all the elements of a tort (for example duty, breach of duty, causation, and damages) are known, or should have been known, since these establish the point at which a plaintiff has a reasonable basis for determining whether or not to bring suit.[84] But requiring knowledge of negligence (the existence of the defendant's legal duty to the plaintiff and breach of that duty) allows some claims to be actionable almost indefinitely and places on the defendant the burden of the plaintiff's failure to find competent advice. Such a situation would arguably defeat the purpose of the FTCA statute of limitations as a statute of repose.[85] The Supreme Court attempted to resolve this controversy in *Kubrick*.[86]

In *Kubrick*, the Court considered an FTCA claim for damages resulting from the plaintiff's deafness, which was allegedly caused by negligent post-surgery treatment of an infection at a Veterans Administration hospital.[87] Although the plaintiff had known of the injury and its probable cause for some time, he had not been able to determine within two years of the surgery whether it was negligently inflicted, despite diligent efforts to pursue this information.[88] Based on this finding of diligence, the trial court held the action timely,[89] and the court of appeals affirmed.[90] The Supreme Court reversed, holding that the statute started to run when the plaintiff had actual knowledge of the

---

[84] Bridgford v. United States, 550 F.2d 978, 981–82 (4th Cir. 1977). *See also* De Witt v. United States, 593 F.2d 276 (7th Cir. 1979); Exnicious v. United States, 563 F.2d 418 (10th Cir. 1977); Jordan v. United States, 503 F.2d 620 (6th Cir. 1974).

[85] *Kubrick*, 444 U.S. at 123–24.

[86] *Id.*

[87] *Id.* at 113–16. Kubrick had sought surgery at a Veterans Administration hospital for an infected femur in April 1968. After surgery, the surgical wound was irrigated with the antibiotic neomycin, and shortly thereafter Kubrick began to experience deafness. In January 1969, an ear specialist diagnosed his condition as bilateral nerve deafness, a permanent condition, and informed him that neomycin treatment was the probable cause. Kubrick was not able to ascertain for some time, however, that this treatment was negligent, and thus he did not file suit for several years. *Id.*

[88] The trial court found that Kubrick had diligently pursued this information. Kubrick v. United States, 435 F. Supp. 166, 185 (E.D. Pa. 1977).

[89] *Id.* Specifically, the court held that actual knowledge of an injury and its cause is sufficient to create a rebuttable presumption that the injury was negligently inflicted. *Id.*

[90] Kubrick v. United States, 581 F.2d 1092 (3d Cir. 1978).

injury and its probable cause.[91] The Court was unwilling to equate "a plaintiff's ignorance of his legal rights [with] his ignorance of the fact of his injury or its cause":[92]

> A plaintiff such as Kubrick, armed with the facts about the harm done to him, can protect himself by seeking advice in the medical and legal community. To excuse him from promptly doing so by postponing the accrual of his claim would undermine the purpose of the limitations statute, which is to require the reasonably diligent presentation of tort claims against the Government . . . . But however or even whether he is advised, the putative malpractice plaintiff must determine within the period of limitations whether to sue or not, which is precisely the judgment other tort claimants must make.[93]

The result of this ruling was stark: Kubrick had already won a substantial award on the merits in the court below.[94]

The result the Court reached in *Kubrick* clearly eliminated any "knowledge of negligence" element from the diligence discovery rule, yet it left several ambiguities. The narrowest reading of the decision triggers the statute of limitations in an FTCA medical malpractice action where actual knowledge of an injury and its probable cause exists.[95] Thus, the Supreme Court implicitly recognized the special situation of medical malpractice claimants by applying a diligence discovery rule of accrual, albeit narrower than the rule previously adopted by most courts under which commencement of the limitation period was deferred until discovery of negligence. However, the Court left unclear whether an objective "knew or should have known" standard applies, or whether the statute commences only when a plaintiff has actual knowledge of an injury and its cause.[96] Further, the Court did not address other situations where application of the diligence discovery rule may be difficult due to

---

[91] *Kubrick*, 444 U.S. at 125. The dissent argued that *Urie* should control and that a plaintiff who was blamelessly ignorant of the cause of action, despite diligently pursuing the necessary information, should be excused from strict application of the two-year period. *Id.* at 127–28 (Stevens, J., dissenting).

[92] *Id.* at 122.

[93] *Id.* at 123–24.

[94] *Kubrick*, 435 F. Supp. at 189.

[95] *See, e.g.*, 3 THE DEPARTMENT OF JUSTICE MANUAL § 4-5.227 (1988) ("[*Kubrick* holds] that a claim accrues within the meaning of Section 2401(b) when the plaintiff knows both the existence and the cause of his/her injury, and not at a later time when he/she also knows that the acts inflicting the injury may constitute medical malpractice").

[96] This ambiguity has led to a variety of applications of the rule. *See* Abney, *For Whom the Statute Tolls: Medical Malpractice Under the Federal Tort Claims Act*, 61 NOTRE DAME L. REV. 696 (1986).

unique factual considerations, such as the government's fraudulent concealment of information. *Kubrick* has been variously applied to such situations; examples of issues and trends in its application are set out below.[97]

Several courts have held that *Kubrick* and the diligence discovery rule generally do not apply during the period of a plaintiff's continuous treatment by the allegedly negligent doctor or hospital.[98] This exception is motivated by a reluctance to impose the ordeal of litigation on a patient receiving treatment, especially when some hope exists that the condition will improve.[99] This special rule seems to conflict with the general tenor of the Supreme Court's discussion,[100] since the plaintiff in such circumstances is temporarily excused from any diligence obligation. Similarly, courts have apparently departed from the *Kubrick* approach where the plaintiff could not mentally comprehend the injury or its cause precisely because of the government's negligence.[101] Situations such as these highlight the difference between a discovery rule that includes an objective "knew or should know" component, and a rule that is purely subjective.

Several other cases present issues that test the parameters of the diligence discovery rule. For instance, in some cases potential plaintiffs have been misled by government doctors into thinking that a certain injury was not negligently inflicted, but rather comprised part of the normal treatment process.[102] Some

---

[97] For a thorough analysis of how *Kubrick* has been applied in medical malpractice cases by each federal circuit court of appeals, see Grasso, *The Statute of Limitations as Applied to Medical Malpractice Actions Brought Under the Federal Tort Claims Act*, 117 MIL. L. REV. 1 (1987). For an analysis of *Kubrick* as it is used in different contexts, see Wagner, *United States v. Kubrick: Scope and Application*, 120 MIL. L. REV. 139 (1988).

[98] *See, e.g.*, Wehrman v. United States, 830 F.2d 1480 (8th Cir. 1987).

[99] *Id.* at 1485.

[100] The Court's discussion indicates that § 2401(b) is meant to be construed narrowly, a situation that would not allow for a baldly equitable exception such as the continuous treatment doctrine. *Kubrick*, 444 U.S. at 117 ("we are not free to construe [§ 2401(b)] so as to defeat its obvious purpose, which is to encourage the prompt presentation of claims"); *see also* Note, *Federal Tort Claims Act*, 14 SUFFOLK U.L. REV. 1428, 1439 (1980).

[101] *See* cases cited *supra* note 6; *but see* Barren v. United States, 839 F.2d 987, 994 & n.2 (3d Cir. 1988) ("[i]t may be that when the negligence of the defendant impairs the ability of a plaintiff to take the necessary measures to file a claim, fairness requires that we relax the [application of the reasonable diligence] rule . . . . That, however, is an issue for Congress . . . [i]nclusion of a plaintiff's mental capacity as a factor to be considered in determining the reasonableness of plaintiff's diligence runs counter to [the] general approach"), *cert. denied*, 109 S.Ct. 79 (1988).

[102] *See, e.g.*, McDonald v. United States, 843 F.2d 247 (6th Cir. 1988); Colleen v. United States, 843 F.2d 329 (9th Cir. 1987).

courts have held that these assurances relieve the plaintiff from compliance with the "diligence" component of the discovery rule, because they prevent a plaintiff from knowing of a particular injury or its cause.

Other courts, however, have upheld the diligence requirement even where the injury itself was expected, and therefore gave the plaintiff no reason to suspect any negligence.[103] In addition, some courts have held that even where a plaintiff has been lulled into the belief that negligence was not involved, *Kubrick* nonetheless requires a search for any potential negligence once the injury and its cause are known.[104] In situations such as these where the injury may have both a negligent and non-negligent cause, discovery of the latter controls for purposes of the discovery rule.

Thus *Kubrick* did not settle the issues concerning when a claim accrues under the FTCA. The factual posture before the Court in that case has been seen by several lower courts as too narrow to have generated broad guidelines applicable to other tort situations.

C. *The Diligence Discovery Test in Practice*

Under *Kubrick*, where a plaintiff actually identifies through due diligence the tortious activity and its origin, the statute of limitations should begin to run at the time of the injury. But determining at what point the plaintiff failed to exercise due diligence is often a difficult task. "The question of what knowledge should put a claimant on notice of the existence of a viable claim is not soluble by any precise formula."[105] The Supreme Court offered guidance in *Kubrick* by mentioning two elements that would necessitate a further investigation by a plaintiff into a possible cause of action: the fact of the injury, and the identity (though not necessarily the governmental capacity) of the tort-

---

[103] *See, e.g.*, Sexton v. United States, 832 F.2d 629 (D.C. Cir. 1987) (parents of child who died in 1968 of complications from experimental radiation treatment for leukemia should have looked into their claim then, even though death was expected; they could not bring an action 10 years later when they discovered that the treatment was improper).

[104] *Id.* at 636.

[105] Waits v. United States, 611 F.2d 550, 552 (5th Cir. 1980) (upholding award to a plaintiff who suffered amputation of his leg, where the Veterans Administration hospital failed to provide him with his records in time to assess the merits of his claim within the two-year limitation period).

feasor—the "what" and "who" of the alleged tort.[106] Therefore, under *Kubrick*, if a plaintiff knows that he was injured by an act, and that it was the defendant who performed the act, the statute of limitations begins to run whether or not the plaintiff knew that the act was illegal.[107]

Where a plaintiff was not clearly in possession of these "critical facts" about the injury and the tortfeasor's identity, the relevant test in determining whether the plaintiff acted with the requisite diligence is the "objective standard of a hypothetical reasonable man."[108] Thus, when a "person of ordinary intelligence [has] knowledge of facts sufficient to suggest" that he has been injured, the statute commences to run, regardless of the plaintiff's ignorance about the tortfeasor's identity.[109]

In addition, while notice of a particular cause of action may be required to trigger the limitation period,[110] courts have indicated in other contexts that general suspicions surrounding a possible cause of action may also be sufficient to require further inquiry.[111] Likewise, the commencement of a related lawsuit

---

[106] *Kubrick*, 444 U.S. at 122.

[107] *Id.* at 113–16. *See also* Liuzzo v. United States, 485 F. Supp. 1274 (E.D. Mich. 1980). In *Liuzzo*, the children of a civil rights worker sued the United States and the FBI for alleged involvement in the murder of the plaintiffs' mother by Klansmen. The plaintiffs alleged that the FBI had prior knowledge, through an informant, of a conspiracy to murder the victim, but they did nothing to prevent it. In applying the *Kubrick* rationale, the court held that the plaintiffs' cause of action was not time-barred because, under those particular circumstances, the cause of action could not have accrued until the plaintiffs had reason to know the tortfeasor's identity. The court said:

> [T]he plaintiffs in this case had no reason to investigate the cause of their mother's death in 1965, because, based on the information known at that time, they believed and were led to believe the cause to be the three Klansmen. Moreover, even if they had had reason to investigate further in 1965, it is not clear that the FBI should reasonably have been a target of their inquiries.

Id. at 1284.

[108] Long v. Abbott Mortgage Corp., 459 F. Supp. 108, 116–17 (D. Conn. 1978) (finding that the plaintiff in a securities fraud action did not exercise due diligence when he failed to investigate the defendants' activities).

[109] Renz v. Beeman, 589 F.2d 735, 751 (2d Cir. 1978) (quoting Sielcken-Schwartz v. American Factors, Ltd., 265 N.Y. 239, 246, 192 N.E. 307, 310 (1934)), *cert. denied*, 444 U.S. 834 (1979). In *Renz* a beneficiary of a family trust sought to impose a constructive trust on 2000 shares of voting preferred stock purchased by a trustee (who was also a beneficiary of the trust) and to remove the trustee. The Second Circuit held that the plaintiffs had the duty to inquire "with diligence" into the activities of the defendant when they received a letter describing the defendant's stock ownership, even though the letter did not describe the source of the ownership. 589 F.2d at 750–51.

[110] *See* Hobson v. Wilson, 737 F.2d 1, 35 (D.C. Cir. 1984).

[111] In Rickel v. Levy, 370 F. Supp. 751 (E.D.N.Y. 1974), the plaintiffs brought a class action suit to seek redress for allegedly fraudulent real estate investment schemes. In the pleadings, the plaintiffs admitted that they had been "suspicious generally" of the defendants' activities more than two years before bringing the suit. The court found that had the plaintiffs been reasonably diligent, they would have conducted further investigations when suspicions were aroused, and the court accordingly dismissed the

may provide the means necessary for investigating possible wrongful activity.[112]

The courts have also inconsistently resolved the question of whether the plaintiff must know of the tortfeasor's government status before the limitation period begins to run. In *Scott v. Casey*,[113] the plaintiffs brought an FTCA suit against the government for injuries resulting from their participation in a medical experiment at a federal prison. The plaintiffs claimed that although the experiments took place in the late 1950's, it was not until the late 1970's that they learned of the federal government's involvement in the testing. The court found that the plaintiffs knew that they had suffered injuries from the test, that the test was conducted at a federal prison hospital, and that they were to receive good-time credit for their participation. The court entered judgment for the defendants, saying it was "hard pressed to believe that, given the foregoing circumstances, the plaintiffs were unaware of the government's involvement in the study," and that armed with this knowledge, they should have sought legal advice.[114] *Scott* demonstrates that where a potential claimant is aware of facts which would alert a reasonable individual that a particular party is involved in certain conduct, the claimant will be held to have known of that party's involvement.

In contrast, *Peck v. United States*[115] provides an example of a case where application of the "diligence discovery" standard did not require dismissal. *Peck* was a civil action against the FBI for an alleged violation of the plaintiff's constitutional

---

suit as time-barred. *Id.* at 756–57; *accord* Drazan v. United States, 762 F.2d 56 (7th Cir. 1985). *See also* Kramer v. Secretary, U.S. Dep't of the Army, 623 F. Supp. 505, 510–11 (E.D.N.Y. 1985).

[112] In Klein v. Shields & Co., 470 F.2d 1344 (2d. Cir. 1972), a brokerage firm had originally brought suit in 1960 alleging that Klein had breached a contract to purchase securities; Klein counterclaimed, alleging that the brokers had failed to deliver the securities. The brokerage firm ceased prosecution of the action in 1961, but in 1971, Klein brought suit on the same grounds and added a claim of securities fraud. The district court dismissed the complaint, holding that the action was time-barred. In affirming the lower court, the Second Circuit held that Klein could have discovered the alleged fraud in 1960 when the possibility of fraud should have been apparent, and had the means to take depositions that "would have provided whatever further insight was necessary." *Id.* at 1346–47.

For other examples in which courts have expected a reasonably diligent potential plaintiff to investigate the possibility of a claim, see Korweck v. Hunt, 646 F. Supp. 953, 958–59 (S.D.N.Y. 1986); Berry Petroleum Co. v. Adams & Peck, 518 F.2d 402, 410 (2d Cir. 1975).

[113] 562 F. Supp. 475 (N.D. Ga. 1983).

[114] *Id.* at 482.

[115] 470 F. Supp. 1003 (S.D.N.Y. 1979).

rights. The plaintiff claimed that, in 1961, the FBI had learned through an informant of an impending assault on the plaintiff, but did not act to prevent it. The informant testified in other trials in 1965, where he revealed that he had been present at the incident, and that he had called the FBI prior to the assault. Newspaper articles, published during the trial, reported the identity of the informant and implied that he might have been involved in the incident.[116] The court said that these facts alone did not reveal that the FBI might have had prior knowledge of the assault, and held that further discovery was necessary to determine if the plaintiff had been on notice of a possible cause of action.[117] Other courts have expressed this thought as a rule that where a "critical element" of a claim is not known to a plaintiff, the statute does not commence to run.[118]

D. *Rationale Underlying the Diligence Discovery Rule*

In *Kubrick*, the plaintiff had actual knowledge of his injury and the identity of the person who had inflicted it.[119] The Court concluded that once the plaintiff knew the "critical facts that he has been hurt and who has inflicted the injury," he knew enough to commence the running of the statute of limitations; the burden was then on the plaintiff to "protect himself by seeking advice in the medical and legal community" to determine the existence of a cause of action.[120]

To resolve the issue before it in *Kubrick*, the Court examined the purposes of the limitations statute and found that the lower court's decision did not serve those purposes. The Court concluded that postponing accrual until a plaintiff became aware that his injury was negligently inflicted "would undermine the purpose of the limitations statute, which is to require the reasonably diligent presentation of tort claims against the Government."[121] This statement focused the analysis on the issue of diligence, but the Court did not address what "reasonable diligence" means and why its exercise is important.

---

[116] *Id.* at 1019–20.
[117] *Id.* at 1020.
[118] *See, e.g.*, Cain v. United States, 643 F. Supp. 175, 179–80 (S.D.N.Y. 1986); Bergman v. United States, 551 F. Supp. 407, 420–22 (W.D. Mich. 1982).
[119] *Kubrick*, 444 U.S. at 118.
[120] *Id.* at 123.
[121] *Id.*

Thus, under *Kubrick* as applied in other decisions, an approach has developed toward statute of limitations problems by which "accrual may be postponed until the plaintiff has or with reasonable diligence should have discovered the critical facts of both his injury and its cause."[122] The diligence discovery rule should not be characterized as an equitable tolling of the statute, notwithstanding the obvious concern with fairness inherent in the diligence discovery rule. One commentator flatly asserted that the FTCA cannot be so tolled,[123] although at least one court has purported to do so.[124] The equitable tolling cases are quite specific in their genesis and limited in scope, and do not offer a broad exception to applying an otherwise appropriate statute of limitations.[125] In *Kubrick* and other cases the courts have instead rationalized the adoption of a diligence discovery rule by stating that it more adequately serves the purposes of the statute than do other rules, and that it leads to fundamentally fair results.

As noted above, while designed to give potential plaintiffs a reasonable time to bring their claims, limitation periods are also designed to "protect defendants and the courts from having to deal with cases in which the search for truth may be seriously impaired by the loss of evidence, whether by death or disappearance of witnesses, fading memories, disappearance of documents, or otherwise."[126] The articulated focus of statutes of limitations, therefore, seems to be on the difficulties inherent in litigating stale claims and on the unfairness to defendants that would result from having to search for evidence that has faded or has been lost due to the passage of time.[127]

---

[122] Barrett v. United States, 689 F.2d 324, 327 (2d Cir. 1982); *see Urie*, 337 U.S. 163; *see also* Schroer v. Chimura, 634 F. Supp. 941, 943 (N.D.N.Y. 1986) (applying *Barrett*).

[123] *See* Abney, *For Whom the Statute Tolls*, 61 NOTRE DAME L. REV. 696, 720–21 (1986) (case law uniform in assuming that FTCA cannot be tolled in this fashion); *see generally* Pascarella v. United States, 582 F. Supp. 790 (D. Conn. 1984); Sangeminio v. Zuckerberg, 454 F. Supp. 206 (E.D.N.Y. 1978).

[124] *See, e.g.*, Bowen v. City of New York, 476 U.S. 467, 479–82 (1986).

[125] Sinclair, *Service of Process: Rethinking the Theory and Procedure of Serving Process Under Federal Rule 4(c)*, 73 VA. L. REV. 1183, 1274–75 (1987).

[126] *Kubrick*, 444 U.S. at 117.

[127] *See also* Walker v. Armco Steel Corp., 446 U.S. 740, 751 (1980) ("[t]he statute of limitations establishes a deadline after which the defendant may legitimately have peace of mind; it also recognizes that after a certain period of time it is unfair to require the defendant to attempt to piece together his defense to an old claim"); Order of R.R. Telegraphers v. Railway Express Agency, 321 U.S. 342, 348–49 (1944) ("[s]tatutes of limitation . . . in their conclusive effects are designed to promote justice by preventing surprises through the revival of claims that have been allowed to slumber until evidence has been lost, memories have faded, and witnesses have disappeared").

This concern with the inequitites defendants suffer when faced with stale claims suggests a presumption behind the statutes of limitations that it is fair to require a plaintiff to bring a case within a certain time regardless of how just a claim may be.[128] If concerns about fading evidence and fairness to defendants were the controlling factors in the analysis of whether or not a limitation period should be strictly applied, then the plaintiff's conduct and ability to determine the critical facts of a case would have little or nothing to do with the accrual of the cause of action. Nonetheless, in some cases concerns about fading evidence and fairness to defendants are subordinated to the nature of the plaintiff's circumstances, and the plaintiff is granted a reprieve from the time bar. The Supreme Court has said that the "policy of repose, designed to protect defendants, is frequently outweighed . . . where the interests of justice require vindication of the plaintiff's rights."[129] Thus, the assertion some courts have flatly made that the FTCA cannot be "equitably tolled"[130] appears questionable.

The shift of focus to concerns about "vindication of the plaintiff's rights"[131] is evident in the development of the "blameless ignorance" principle by which the limitation period does not commence until the plaintiff knew or should have known the critical facts regarding the cause of action.[132] The genesis of this

---

[128] *See Order of R.R. Telegraphers*, 321 U.S. at 348–49.

[129] Burnett v. New York Cent. R.R. Co., 380 U.S. 424, 428 (1965). *Burnett* held that the statute of limitations under the Federal Employers' Liability Act ("FELA") was tolled when the plaintiff filed suit in state court. The Court said that the plaintiff's failure to file in federal court was "not because he was disinterested, but solely because he felt that his state action was sufficient. Respondent could not have relied upon the policy of repose embodied in the limitation statute, for it was aware that petitioner was actively pursuing his FELA remedy." *Id.* at 429–30. Thus, the Court found that the policy of repose was outweighed by the plaintiff's right to compensation. Interestingly, in Barren v. United States, 839 F.2d 987, 995 (3d Cir. 1988), one dissenting judge apparently thought that if the court focused on the plaintiff's circumstances, the policy of repose would take care of itself. He rejected the government's claim that an equitable approach to determining accrual "is 'unworkable as a practical matter' because it 'does not indicate how long accrual should be delayed.'" *Id.* at 999 (Becker, J., dissenting). He said that this should not concern the court because "a plaintiff who brings suit after too long a period will have trouble adducing evidence and proving his claim." *Id.* at 1000. Thus, this jurist considered the two-year period to be a useful—if somewhat arbitrary—deadline for determining when a claim has become stale, and focused instead on the fairness to the plaintiff; he thereby assumed that the tendency of evidence to disappear over time would operate as a practical check on bringing stale claims.

[130] *See Pascarella*, 582 F. Supp. 790; Sangeminio v. Zuckerberg, 454 F. Supp. 206 (E.D.N.Y. 1978); *see generally* Abney, *supra* note 96, at 720–21.

[131] *Burnett*, 380 U.S. at 428.

[132] *See Kubrick*, 444 U.S. at 120 n.7.

principle can be traced to *Urie v. Thompson*,[133] in which the Supreme Court tolled the statute of limitations under the Federal Employers' Liability Act in light of the plaintiff's "blameless ignorance" of the "inherently unknowable" facts of his work-related injury.[134] The Court said:

> [We did] not think the humane legislative plan intended such consequences to attach to blameless ignorance. Nor do we think those consequences can be reconciled with the traditional purposes of statutes of limitations, which conventionally require the assertion of claims within a specified period of time after notice of the invasion of legal rights.[135]

By the time the Court decided *Kubrick*, the federal courts had generally recognized that "any plaintiff who is blamelessly ignorant of the existence or cause of his injury should be accorded the benefits of the more liberal accrual standard."[136] This approach colored the Court's interpretation of the FTCA, notwithstanding that the FTCA is a limited waiver of sovereign immunity and that courts were cautioned not to "extend the waiver beyond that which Congress intended."[137] While this reminder should temper those courts eager to extend the statute of limitations period on equitable grounds,[138] section 2401(b)'s imprecise language necessitates construction by the courts in all but the most straightforward cases. It is not simply a matter of deciding whether to extend the waiver; the courts must define what section 2401(b) means.[139] In that process, recourse to general principles of interpretation, including equitable considerations, is inescapable. While courts must be careful not to take it upon themselves to grant relief beyond that intended by Congress, they must also be mindful of the FTCA's remedial purpose.

Thus "vindication of the plaintiff's rights"[140] does not become any less important because the FTCA is a limited waiver of

---

[133] 337 U.S. 163 (1949).
[134] *Id.* at 169–71.
[135] *Id.* at 170.
[136] *Barrett*, 689 F.2d at 327.
[137] *Kubrick*, 444 U.S. at 117–18.
[138] *See* Houston v. United States Postal Serv., 823 F.2d 896 (5th Cir. 1987), *cert. denied*, 485 U.S. 1006 (1988), in which the court said that "[e]quitable considerations that may waive or toll limitations periods in litigation between private parties do not have the same effect when suit is brought against the sovereign. This is because under the doctrine of sovereign immunity the government's exposure to liability can be no greater than it permits." *Id.* at 902 (citation omitted).
[139] *See supra* notes 43–44 and accompanying text.
[140] Burnett v. New York Cent. R.R. Co., 380 U.S. 424, 428 (1965).

sovereign immunity. This is implicit in *Kubrick*, where the Supreme Court acknowledged that there are certain "critical facts" that a plaintiff must possess before the limitation period begins to run.[141] Indeed, the Court's statement that the "purpose of the limitations statute" is to require "the *reasonably diligent* presentation of tort claims"[142] incorporated this fairness principle. Although the Court reaffirmed a policy of protecting defendants and the courts from stale claims,[143] it focused on whether or not it would be fair to the plaintiff to toll the statute of limitations until the plaintiff knew that his injury was due to the defendant's negligence.

E. *Burdens and Motion Practice Under a Diligence Regime*

Many limitation issues are raised either in motions brought under Rule 12 of the Federal Rules of Civil Procedure against a plaintiff's pleadings, or in summary judgment motions brought under Rule 56. Using the more nuanced diligence discovery approach to deciding these issues may appear to reduce the likelihood of efficiently disposing of cases on these motions. As discussed below, this is not necessarily so. More importantly, however, the diligence discovery approach may allow some cases to remain on the docket that in fairness deserve to be heard, but which would otherwise be summarily dismissed on limitation grounds.

Rule 12 limitation motions will lie in some FTCA cases even under a broad diligence discovery approach. Where the cases involve obvious physical torts,[144] and the complaint on its face reveals that the defendant acted outside the statutory period,

---

[141] The Court said:

> We are unconvinced that for statute of limitations purposes a plaintiff's ignorance of his legal rights and his ignorance of the fact of his injury or its cause should receive identical treatment. That he has been injured in fact may be unknown or unknowable until the injury manifests itself; and the facts about causation may be in the control of the putative defendant, unavailable to the plaintiff or at least very difficult to obtain. The prospect is not so bleak for a plaintiff in possession of the critical facts that he has been hurt and who has inflicted the injury.

Kubrick, 444 U.S. at 122.

[142] *Id.* at 123 (emphasis added).

[143] *Id.* at 117.

[144] *See supra* text accompanying notes 70–75.

motions addressed to the pleadings will still be feasible.[145] If, however, the plaintiff then responds with an affidavit suggesting that the plaintiff did not know and could not reasonably have known of the wrong until a later date, which is within the statutory period, a motion to dismiss will be converted to a summary judgment notice by operation of law.[146]

Summary judgment may also be granted on limitation grounds, despite the application of the diligence discovery rule to the plaintiff's conduct.[147] On a limitation motion, as in other summary judgment applications under Rule 56, the burden in the first instance is on the moving party to demonstrate that there is no genuine issue as to any material fact and that it is entitled to judgment as a matter of law.[148] Ambiguities must be resolved against the moving party[149] and uncertainties about material facts will defeat the motion.[150] An additional doctrine applies in at least some circuits, however, so that after the defendant has established that conduct at issue is beyond the normal reach of the statutory period, the burden shifts to a plaintiff who seeks to pursue claims arising out of events going back more than two years "to prove [his or her] diligence as the prerequisite to allowing equitable tolling of the limitations statute."[151] Nonetheless, in a series of 1986 decisions[152] the Supreme Court reaffirmed the longstanding principle that the presence of issues of fact precludes summary judgment on such limitation questions under the Act.[153]

---

[145] *See, e.g.*, Zavala *ex rel.* Ruiz v. United States, 876 F.2d 780 (9th Cir. 1989) (medical malpractice action dismissed under Rule 12(b)(1) for lack of subject matter jurisdiction where administrative claim was filed more than four years after cause of action accrued); Barnhart v. United States, 884 F.2d 295, 296 (7th Cir. 1989) (medical malpractice action under the FTCA in which the defendant's original summary judgment motion was converted to a motion to dismiss under Rule 12(b)(1); such preliminary matters most appropriately treated as motions to dismiss), *cert. denied*, 110 S.Ct. 2561 (1990).

[146] FED. R. CIV. P. 12. *See* Snyder v. United States, 717 F.2d 1193, 1194 (8th Cir. 1983).

[147] *See, e.g.*, Dulaine v. United States, 371 F.2d 824 (9th Cir. 1967), *cert. denied*, 387 U.S. 920 (1967).

[148] FED. R. CIV. P. 56. *See, e.g.*, Hamilton v. Smith, 773 F.2d 461, 466 (2d Cir. 1985).

[149] *Hamilton*, 773 F.2d at 466.

[150] *Id.*

[151] Freschi v. Grand Coal Venture, 767 F.2d 1041, 1047 (2d Cir. 1985) (citing City of Detroit v. Grinnell Corp., 495 F.2d 448, 461 (2d Cir. 1974)), *cert. granted and judgment vacated*, 478 U.S. 1015 (1986); Long v. Abbott Mortgage Corp., 459 F. Supp. 108, 116 (D. Conn. 1978).

[152] *See* Celotex Corp. v. Catrett, 477 U.S. 317 (1986), *cert. denied*, 484 U.S. 1066 (1988); Anderson v. Liberty Lobby, Inc., 106 S.Ct. 2505 (1986). Although these important decisions made it easier to bring a successful summary judgment motion and dispose of baseless claims before trial, they did not resolve the more particular questions about the appropriateness of a summary judgment based on a limitation issue.

[153] *See, e.g.*, Cain v. United States, 643 F. Supp. 175, 179–80 (S.D.N.Y. 1986);

Several decisions refer to the "jurisdictional" nature of the limitation defense.[154] However, the limitation provision should not be construed as self-effectuating; the mere raising of a motion on this ground does not deprive the district court of subject matter jurisdiction over the cause of action. As the above-mentioned summary judgment cases make clear, fact issues often control when the viability of the suit is not evident from the face of the pleadings, and the court has the power to retain the suit while supervising the motion practice in order to streamline the issues. Indeed, there are numerous reported decisions in which the trial court has elected to reserve the limitation question until trial.[155] Judges would otherwise risk deciding pre-trial motions, including issues of what "should have been known," without requisite factual background.[156] Some cases will remain in which pleadings and affidavits cannot give the court sufficient depth of understanding to allow comfortable disposition of limitation questions on motion. The defense is not foregone in such situations, of course, and is simply presented at trial[157] as a decision for the trier of fact.[158]

Cordaro v. Lusardi, 354 F. Supp. 1147 (S.D.N.Y. 1973), *aff'd without opinion*, 513 F.2d 624 (2d Cir. 1975). Some decisions have suggested that only in "extreme circumstances" is summary judgment appropriate on limitation issues. *See* Freschi, 583 F. Supp. at 785.

[154] *See, e.g.*, cases cited *infra* notes 162–163.

[155] *See, e.g.*, Oslund v. United States, 701 F. Supp. 710 (D. Minn. 1988) (where the plaintiff presented strong evidence that he had been severely dysfunctional for years and had only recently gained the capacity to understand the cause of his injury, question of whether his action was barred by the statute of limitations was for the jury).

[156] *See generally* Zavala *ex rel.* Ruiz v. United States, 876 F.2d 780 (9th Cir. 1989) (court dismissed medical malpractice case pursuant to Rule 12(b)(1) but only after finding facts from evidence presented, and determining that abandonment by the injured infant's father did not toll the period of limitations since he had knowledge of the child's condition and responsibility for the child).

See also Herrera-Diaz v. United States, 845 F.2d 1534, 1541 (9th Cir. 1988) (Norris, J., dissenting), *cert. denied*, 109 S. Ct. 306 (1988). The dissent stated:

> [D]iscovery of this technical cause simply cannot be deemed sufficient knowledge of cause to trigger the statute of limitations. Without some suggestion that a human agent might have contributed to her son's injury, [the mother] cannot be faulted for not inquiring further about the cause of [the boy's] cerebral palsy. In other words, the record as it now stands fails to establish that [the mother] had sufficient knowledge about who inflicted her son's injury.

Id. at 1541.

[157] *See, e.g.*, Oberlin v. United States, 727 F. Supp. 946, 949 (D. Pa. 1989) (issue of fact as to when claim accrued in medical malpractice action precluded summary judgment for the government); Monday v. United States, 688 F. Supp. 788, 791 (D. Me. 1988) (genuine issues of fact as to whether parents knew of their son's injuries within the meaning of *Kubrick*, and whether they knew of the probable cause of his injuries before diagnosis, precluded summary judgment).

[158] *See Oslund*, 701 F. Supp. at 712 ("[t]he ultimate resolution of whether the statute should be tolled must wait until trial. At trial both sides may present their conflicting evidence, and the trier of fact will decide the issue").

The jurisdictional nature of the defense, therefore, raises one last question: whether failure to raise the defense waives it, or whether the defendant may raise the limitation period as a bar at any time, without having included it in the pleadings. Many cases in contexts other than the FTCA have held the statute of limitations to be an affirmative defense under Rule 8,[159] which a defendant must assert or waive.[160] And where the United States is a plaintiff, a defendant who fails to assert the statute of limitations defense punctually waives that defense.[161] Where the United States is a defendant, however, older decisions do tend to discuss the jurisdictional nature of the defense and usually conclude that the action must be dismissed even though the government did not plead the statute of limitations.[162] In one of the few cases to discuss this rule under the FTCA, decided

---

[159] FED. R. CIV. P. 8(c) ("[i]n pleading to a preceding pleading, a party shall set forth affirmatively . . . [the] statute of limitations").

[160] *See, e.g.*, Davis v. Bryan, 810 F.2d 42 (2d Cir. 1987) (in civil rights action against the government brought by a prisoner, district court judge erred in granting summary judgment to appellees on statute of limitation grounds because appellees had not pleaded the defense, but the court had raised the statute of limitations issue *sua sponte*); Banks v. Chesapeake and Potomac Tel. Co., 802 F.2d 1416, 1427 (D.C. Cir. 1986) (in race and sex discrimination action against former employer, "reliance on statute of limitations is an affirmative defense and is waived if a party does not raise it in a timely fashion" (citations omitted)); Chapman v. Orange Rice Milling Co., 747 F.2d 981 (5th Cir. 1984) (in breach of contract action brought by lessor, court held that the statute of limitations, an affirmative defense, must be specifically pled or is waived).

But see Expertise, Inc. v. Aetna Fin. Co., 810 F.2d 968 (10th Cir. 1987) (in breach of contract action, statute of limitations was not waived though it was not raised in the defendant's answer because it was included in the pretrial order); Rivera v. Anaya, 726 F.2d 564, 565 (9th Cir. 1984) (in action brought by migrant farm workers against employer for violating government regulations, the defendant's "failure to raise the statute of limitations as a defense in response to the first pleading did not serve to waive his right to raise it later [in a summary judgment motion] absent prejudice to the plaintiffs").

[161] United States v. Ward, 618 F. Supp. 884, 901 (D.C.N.C. 1985) (in action by the United States to recover expenses incurred in cleaning up sites at which oil containing polychlorinated biphenyls had been dumped, the defendants waived statute of limitations defense as to counterclaims against them by failing to include that defense in reply to counterclaims and failing to seek to amend reply to add that defense); United States v. Eytcheson, 237 F. Supp. 371 (D. Mont. 1965) (in action by the United States for damage caused to property by forest fire, statute of limitations defense waived where not raised in answer).

[162] *See* Isthmian S.S. Co. v. United States, 191 F. Supp. 338 (S.D.N.Y. 1961) (in action against the United States in admiralty to recover sums deducted from freight demurrage charges, statutes of limitation deemed jurisdictional, and therefore applicability of six-year limitation had to be considered, even though it was not within the issues formulated by pre-trial stipulation or pleaded as a basis of jurisdiction), *aff'd*, 302 F.2d 69 (2d Cir. 1962); Werner v. United States, 10 F.R.D. 245 (S.D. Cal. 1950) (in action against the United States for reformation of a lease, where the claim was barred because the action was not brought within six years from the date the cause of action accrued, the action would be dismissed for want of jurisdiction even though the government had not pleaded the statute barring the claim), *aff'd*, 188 F.2d 266 (9th Cir. 1951).

shortly after the limitation provision took its present form, the court held that where an action against the United States under the FTCA is barred by the limitation provision, the court must dismiss the action regardless of whether the limitation is pleaded.[163] Nevertheless, given the increased rigor with which diligence notions are being applied to government litigation today,[164] in the future the statute of limitations defense may be waivable even under the FTCA.

III. BEYOND THE *KUBRICK* PARADIGM

However useful courts may have found *Kubrick* in deciding cases with similar facts,[165] they continue to apply the diligence discovery rule to dissimilar cases in an ad hoc manner. This reflects the varying levels of comfort courts experience in introducing equitable principles into the analysis. Thus courts have sometimes concluded that they are powerless to use equity to prevent strict application of the limitation period and thereby avoid a harsh result.[166]

Just after *Kubrick* was decided, in *Steele v. United States*[167] the Seventh Circuit refused to apply a diligence discovery rule to a plaintiff who alleged that the Federal Aviation Administration negligently left electric current on a transformer from which the plaintiff received an electric shock.[168] The court cited a number of reasons why diligence discovery may uniquely apply to the medical malpractice context: the injury involved may go unnoticed, its origin may be especially difficult to learn, and the causal connection may be especially difficult to ascertain; the standard of care in such cases is therefore difficult to judge.[169] The court said that if the diligence discovery rule applies outside the medical malpractice context at all, then the threshold of

---

[163] DeBonis v. United States, 103 F. Supp. 123 (W.D. Pa. 1952).

[164] Paetz v. United States, 795 F.2d 1533 (11th Cir. 1986) (in age discrimination suit by former civil service employee, the government's failure to assert statute of limitations as an affirmative defense in pleadings constituted a waiver of the defense), *cert. denied*, 110 S.Ct. 74 (1989).

[165] *See, e.g.*, Fernandez v. United States, 673 F.2d 269, 270–72 (9th Cir. 1982).

[166] *See* Barren v. United States, 839 F.2d 987, 996–97 (3d Cir. 1988) (Becker, J., dissenting), *cert. denied*, 488 U.S. 827 (1988). Judge Becker in dissent spoke out against what he saw as "the obvious injustice" of a strict interpretation of the *Kubrick* rule without taking into account "[b]asic principles of equity." *Id.*

[167] 599 F.2d 823 (7th Cir. 1979).

[168] *Id.* at 824–28.

[169] *Id.* at 828.

discovery should be lowered. The plaintiff should only need to know, actually or constructively, "the last essential element of the tort, *i.e.*, the damage" to commence the running of the statute of limitations.[170]

Similarly, the Fifth Circuit described the diligence discovery rule applied in *Kubrick* as "the medical malpractice accrual test,"[171] finding that a different test "similar to, but distinct from, the medical malpractice accrual test" applies outside the malpractice context.[172] This was so because the *Kubrick* test developed "to protect those who suffered damage arising out of both a specialized area, medicine, and a unique relationship, doctor-patient."[173]

Yet a number of courts have decided that the *Kubrick* diligence discovery rule is not limited to the medical malpractice context. Thus, "any plaintiff who is blamelessly ignorant of the existence or cause of his injury shall be accorded the benefits of the discovery rule . . . . [T]he rule was not created in a medical malpractice context and is not limited to such cases."[174]

In *Guccione v. United States*,[175] the court applied a *Kubrick* diligence discovery standard to a plaintiff alleging injuries sustained during the course of the FBI's ABSCAM investigation. The court said, "[t]his special rule of accrual may . . . be applied where it has been shown that the plaintiff was blamelessly ignorant of his claim due to the government's deliberate concealment of its facts."[176]

---

[170] *Id.*

[171] Ware v. United States, 626 F.2d 1278, 1284 n.4 (5th Cir. 1980).

[172] *Id.*

[173] *Id.* at 1284 n.4. The court held that a cause of action for destruction of cattle accrues when "the injury coincides with the negligent act and some damage is discernable at the time." *Id.* at 1284.

[174] Stoleson v. United States, 629 F.2d 1265, 1269 (7th Cir. 1980). In Orlikow v. United States, 682 F. Supp. 77 (D.D.C. 1988), the plaintiff alleged that the Central Intelligence Agency was negligent in secretly funding a doctor who experimented on unwitting human subjects. The court said that "where the 'what' element and the 'who' element is missing [sic] from the puzzle, a claim against the government should be tolled if plaintiff had expended due diligence to uncover these facts." *Id.* at 84. In Bush v. United States, 823 F.2d 909 (5th Cir. 1987), the court applied the diligence discovery rule to a plaintiff who alleged that the State Department failed to release available medical records, which allegedly resulted in his inability to obtain adequate medical treatment. *Id.* at 910–11.

[175] 670 F. Supp. 527 (S.D.N.Y. 1987), *aff'd on other grounds*, 847 F.2d 1031 (2d Cir. 1988), *cert. denied*, 110 S.Ct. 719 (1990).

[176] *Id.* at 536; *see also* Barrett v. United States, 689 F.2d 324, 327 (2d Cir. 1982), *cert. denied*, 462 U.S. 1131 (1983), where the court stated:

> The diligence discovery rule of accrual is not often applied outside the medical malpractice area, but may be appropriate in non-malpractice cases, where plaintiffs face comparable problems in discerning the fact and cause of their

Thus, when faced with facts that do not fall within the *Kubrick* paradigm, some courts have continued to consider the appropriate application of equitable concerns in determining when a cause of action accrues under the FTCA. A clear consensus on the permissible scope of equity has yet to emerge. The following discussion highlights three categories of cases outside medical malpractice where courts have applied diligence discovery principles.

A. *Active Concealment*

Courts have long recognized that cases in which a defendant is alleged to have deliberately concealed the facts relating to the alleged tortious actions present difficult limitation issues.[177] In suits under the FTCA, such issues have arisen in actions involving law enforcement investigations,[178] medical tests,[179] and other circumstances.[180]

Of course, it is not always easy to determine whether conduct amounts to fraud or deliberate concealment. Mere nondisclosure does not necessarily constitute concealment; it is said that the government has no obligation to discover its agents' torts and

---

> injuries. Thus, any plaintiff who is blamelessly ignorant of the existence or cause of his injury should be accorded the benefits of the more liberal accrual standard.

Id. at 327 (citations omitted).

[177] *See generally* Dawson, *Estoppel and Statutes of Limitation*, 34 MICH. L. REV. 1 (1935); Dawson, *Undiscovered Fraud and Statutes of Limitation*, 31 MICH. L. REV. 591 (1933); Dawson, *Fraudulent Concealment and Statutes of Limitation*, 31 MICH. L. REV. 875 (1933).

[178] *See, e.g.*, Hobson v. Wilson, 737 F.2d 1 (D.C. Cir. 1984), *cert. denied*, 470 U.S. 1084 (1985); Guccione v. United States, 670 F. Supp. 527 (S.D.N.Y. 1987); Socialist Workers Party v. Att'y Gen. of United States, 642 F. Supp. 1357 (S.D.N.Y. 1986); Bergman v. United States, 551 F. Supp. 407 (W.D. Mich. 1982).

[179] *See, e.g.*, Barrett v. United States, 689 F.2d 324 (2d Cir. 1982), *aff'd*, 853 F.2d 124, *cert. denied*, 462 U.S. 1131 (1983); Orlikow v. United States, 682 F. Supp. 77 (D.D.C. 1988); Scott v. Casey, 562 F. Supp. 475 (N.D. Ga. 1983).

[180] *See, e.g.*, Gibson v. United States, 781 F.2d 1334 (9th Cir. 1986) (FBI allegedly burned down the plaintiff's garage to conceal investigation); Diminnie v. United States, 728 F.2d 301 (6th Cir. 1984) (A plaintiff previously convicted of extortion for sending anonymous threats that federal buildings would be blown up sued a federal agent for allegedly concealing the fact that he, rather than the plaintiff, had sent the threats), *cert. denied*, 469 U.S. 842 (1984); Leftridge v. United States, 612 F. Supp. 631 (W.D. Mo. 1985) (Occupational Safety and Health Administration allegedly failed to discover hazardous conditions and concealed information about its investigation).

publish the facts to prospective plaintiffs.[181] In addition, if a plaintiff has actual notice of a potential claim, the period begins to run despite the fact that the defendant may have attempted to conceal the cause of action.[182]

Several examples illustrate these principles. For instance, cases involving allegations of impermissible law enforcement investigations or techniques, such as physical or electronic surveillance or surreptitious entries into the premises of an investigative target, by their nature concern acts that are intended to remain unknown to the person or entity being investigated.[183] In such circumstances, obvious injustices result if the claim is deemed to have accrued at the time of the plaintiff's injury. Where a plaintiff had no means of discovering the critical facts on which to base a claim within the limitation period, an FTCA remedy is illusory. Additionally, unless the accrual test factors in concealment efforts, those engaged in such activities would have a strong incentive to conceal their actions in the hope that they will escape liability by the mere passage of time. Congress could not have intended for the FTCA to be unavailable to those injured by inherently undiscoverable, wrongful governmental activities. Thus, application of the traditional accrual test is not calculated to allow proper vindication of rights when the nature of the wrongful conduct is intentionally concealed by the actor.

Recognizing this, courts have applied the diligence discovery approach in such cases,[184] thereby avoiding the inequity of either allowing the wrong to go unredressed because technically the limitation period expired before discovery, or allowing the plain-

---

[181] *See* Davis v. United States, 642 F.2d 328 (9th Cir. 1981), *cert. denied*, 455 U.S. 919 (1982); *cf.* Pitts v. Unarco Indus., 712 F.2d 276 (7th Cir. 1983) (passive silence is insufficient to trigger the fraudulent concealment doctrine unless the defendant has a fiduciary relationship with the putative plaintiff), *cert. denied*, 464 U.S. 1003 (1983). *See also* Shock v. United States, 689 F. Supp. 1424 (D. Md. 1988) (failure to disclose voluntarily surgeon's negligence insufficient to make out case of fraudulent concealment).

[182] *See* Hobson v. Wilson, 737 F.2d 1, 35 & n.107 (D.C. Cir. 1984); Tinker v. Abrams, 640 F. Supp. 229, 232–33 (S.D.N.Y. 1986).

[183] *See generally Hobson*, 737 F.2d at 32–35 (reviewing older cases and adhering to the view previously adopted in that circuit that in "self-concealing wrongs" the burden is on the defendant to come forward with proof that the plaintiff could have discovered the cause of action in the exercise of due diligence); *see also* Peck v. United States, 470 F. Supp. 1003, 1018–19 (S.D.N.Y. 1979) (in civil action against the government, in part under the FTCA, for alleged violation of the plaintiff's constitutional rights by the FBI, due diligence rule applied where the plaintiff alleged that the defendant had engaged in covert acts).

[184] *See Barrett*, 689 F.2d at 327.

tiff to delay until receiving "actual notice" of the actionable conduct.

As stated by the Second Circuit in a seminal intentional concealment case, "*every federal statute of limitations* . . . [includes] the equitable doctrine that in case of defendant's fraud or deliberate concealment of material facts relating to his wrongdoing, time does not begin to run until plaintiff discovers, or by reasonable diligence could have discovered, the basis of the lawsuit."[185]

But in some cases not coming within the FTCA (such as the securities fraud cases discussed immediately below), courts of appeals only consider the time to run from actual, not constructive, discovery of the tortious conduct. As the discussion below suggests, while this approach may work in those other contexts,[186] a closer examination of the cases involving alleged deliberate concealment reveals that the "actual discovery" rule is not an appropriate test under the FTCA.

The leading case adopting the actual discovery test for accrual of causes of action in the active concealment circumstance is the Seventh Circuit's decision in *Tomera v. Galt*,[187] a non-FTCA case. In *Tomera*, the plaintiff brought an action for alleged securities fraud, claiming that the defendants had taken positive steps to prevent the plaintiff from discovering illegal activity.[188] The Seventh Circuit held that while the due diligence standard for accrual would apply to fraudulent behavior that went "undiscovered even though the defendant after commission of the wrong does nothing to conceal it,"[189] if "the defendant has taken

[185] *Id.* (quoting Fitzgerald v. Seamans, 553 F.2d 220, 228 (D.C. Cir. 1977) (emphasis added)); *cf.* Crown Coat Front Co. v. United States, 275 F. Supp. 10 (S.D.N.Y. 1967) (distinguishing between "'fraudulent concealment' and failure to ascertain the existence of a cause of action through exercise of due diligence"), *aff'd*, 395 F.2d 160 (2d Cir. 1968), *cert. denied*, 393 U.S. 853 (1968). *See also* Dyniewicz v. United States, 742 F.2d 484 (9th Cir. 1984) ("[t]here are no grounds for tolling the statute of limitations based simply on the *Government's* knowledge of its own wrongdoing absent fraudulent concealment, or other forms of conduct that may be recognized as grounds for equitable tolling of the statute" [citation omitted]) (emphasis in original); Hammond v. United States, 388 F. Supp. 928 (E.D.N.Y. 1975) (concealment by the government is no bar to the application of the limitations period).

[186] *See, e.g.*, Robertson v. Seidman & Seidman, 609 F.2d 583 (2d Cir. 1979); Sperry v. Barggren, 523 F.2d 708 (7th Cir. 1975); Tomera v. Galt, 511 F.2d 504 (7th Cir. 1975).

[187] 511 F.2d 504 (7th Cir. 1975). *See also Sperry*, 523 F.2d at 711 (in securities fraud case, "[s]hould active concealment be found, then the statute [of limitations] is tolled until actual discovery"); McConnell v. Frank Howard Allen & Co., 574 F. Supp. 781, 787–88 (N.D. Cal. 1983) (in securities fraud case, action with allegations of active concealment accrued on the date of actual discovery of the fraudulent conduct).

[188] *Tomera*, 511 F.2d at 509.

[189] *Id.* at 510.

positive steps after commission of the fraud to keep it concealed," the limitations period would be tolled "until actual discovery by the plaintiff."[190]

Robertson v. Seidman & Seidman,[191] a widely cited Second Circuit decision, is generally read to adopt the actual discovery rule announced in *Tomera*.[192] In *Robertson*, a corporation alleged that the defendant, a certified public accounting firm, prepared false and misleading financial documents in an effort to further a conspiracy to defraud the corporation, thus violating the antifraud provisions of the Securities Exchange Act.[193] The district court determined that the plaintiff's cause of action accrued from the time when it could have been alerted to the existence of fraud through the exercise of "reasonable diligence." The court granted the defendant's summary judgment motion and dismissed the complaint as time-barred.[194]

On appeal, the Second Circuit reversed and remanded.[195] The court held that "[u]nder the federal equitable tolling doctrine, the active concealment of fraudulent conduct tolls the statute of limitations in favor of the defrauded party until such time as he actually knew of the fraudulent conduct of the opposing party."[196] The court ruled that should active concealment be found, the statute of limitations would be tolled until actual discovery, and it remanded the case to the district court for a

---

[190] *Id.* The court quotes at length from a Pennsylvania case, Smith v. Blanchley, 198 Pa. 173, 47 A. 985 (1901), in announcing its actual discovery test. *Tomera*, 511 F.2d at 510. Such a reference is incongruous. *Smith* was an action based on an alleged fraudulent scheme to receive a large sum of money from the plaintiff in return for preventing the institution of a criminal prosecution against the plaintiff and his family. *Smith*, 198 Pa. at 174, 47 A. at 985. The Pennsylvania Supreme Court addressed the question of whether alleged fraudulent concealment would toll the statute of limitations. While the court held that allegations of fraudulent concealment would toll the statute, it applied a due diligence standard to the plaintiff's conduct and held that the action was time-barred:

> The gradual leaking out of the circumstances, and the gossip and suspicions of others, started an investigation by plaintiffs, which the most ordinary prudence would have prompted at the beginning, and which would then have either foiled the scheme or led to its discovery, and the trial of this action while all the witnesses were alive and the matters fresh in their memories. As it is now, the evidence is so meager that one jury has disagreed upon it, and another has decided it on oath against oath with very little collateral evidence to help out either—*an illustration of the very evil the statute of limitations was intended to prevent.*

Id. at 180, 47 A. at 987 (emphasis added).

[191] 609 F.2d 583 (2d Cir. 1979).

[192] *See, e.g.*, Baskin v. Hawley, 807 F.2d 1120, 1132 (2d Cir. 1986).

[193] *Id.* at 585.

[194] *Id.* at 586.

[195] *Id.* at 594.

[196] *Id.* at 593.

determination as to "whether there was sufficient concealment . . . to invoke the federal equitable tolling doctrine."[197]

Apart from its limited application to securities fraud litigation, the actual discovery rule has been explicitly rejected in a number of opinions in other circuit courts of appeals.[198] Instead courts have applied the due diligence test to cases of alleged active concealment. For example, *Keating v. Carey*,[199] a non-FTCA Second Circuit decision, involved alleged constitutional violations by the State of New York in terminating the employment of a civil servant. The district court had granted the defendant's summary judgment motion on statute of limitation grounds, and the court of appeals reversed. The court held that "[u]nder federal law, when the defendant fraudulently conceals the wrong, the time does not begin running until the plaintiff discovers, or by the exercise of reasonable diligence should have discovered, the cause of action."[200] The case was remanded "to determine the facts regarding [the plaintiff's] delay in bringing this action."[201]

In *Barrett v. United States*,[202] an action under the FTCA for alleged negligence by the United States in a chemical warfare experiment, the plaintiff's estate claimed that the government had conspired to conceal the facts surrounding the plaintiff's death.[203] The court of appeals, in reversing the lower court's grant of summary judgment to the defendant, determined that if the plaintiff's allegations that the Army Chemical Corps actively covered up certain facts pertinent to the experiment were

---

[197] *Id.*

[198] *See* Campbell v. Upjohn Co., 676 F.2d 1122, 1128 (6th Cir. 1982), where the Sixth Circuit "declined to formulate a separate rule for cases involving active concealment by the defendant," holding that allegations that the defendant had concealed terms of an allegedly fraudulent merger agreement would "not exempt the plaintiff from the requirement of diligence in pleading the federal equitable tolling doctrine of fraudulent concealment"; State of Ohio v. Peterson, 651 F.2d 687, 694–95 (10th Cir. 1981) (Tenth Circuit affirmed lower court's application of due diligence standard to action by state against law firm for alleged securities fraud, in part because "we see no reason why an act of concealment by defendant should excuse plaintiff from his obligation of diligence which he owes the court as well as his adversaries"), *cert. denied*, 454 U.S. 895 (1981); *In re* Beef Indus. Antitrust Litig., 600 F.2d 1148, 1169–1170 n.27 (5th Cir. 1979) (in antitrust action, "the running of statute of limitations would not have been tolled even had the defendants affirmatively concealed the alleged scheme"), *cert. denied*, 449 U.S. 905 (1980).

[199] 706 F.2d 377 (2d Cir. 1983).

[200] *Id.* at 382. Interestingly, the court cited *Tomera*, discussed *supra* notes 187–190, for this proposition. *Keating*, 706 F.2d at 382.

[201] *Id.* at 388.

[202] 689 F.2d 324 (2d Cir. 1982), *cert. denied*, 462 U.S. 1131 (1982).

[203] *Id.* at 326–27.

true, such action "would constitute deliberate concealment of material facts relating to the Government's wrongdoing and would trigger application of the diligence-discovery accrual standard."[204] The Second Circuit remanded the case for a determination of when the plaintiff "should have discovered the critical facts relating to the cause" of the decedent's death.[205]

Courts have also used the diligence discovery standard for FTCA cases involving alleged illegal law enforcement activities. In *Guccione v. United States*,[206] the district court applied the diligence discovery standard in determining when a cause of action accrued for a plaintiff who alleged that he sustained injuries during the course of the FBI's ABSCAM investigation. The court said that "[t]his special rule of accrual may . . . be applied where it has been shown that the plaintiff was blamelessly ignorant of his claim due to the government's deliberate concealment of its facts."[207]

The diligence discovery rule better coheres with Supreme Court jurisprudence on the deliberate concealment issue than does the actual discovery rule. A series of cases beginning in 1874 with *Bailey v. Glover*,[208] demonstrates the origins of the diligence discovery rule in the Supreme Court's treatment of fraudulent concealment. In *Bailey*, a case brought for alleged bankruptcy fraud, the Supreme Court stated:

> [W]e hold that when there has been no negligence or laches on the part of a plaintiff in coming to the knowledge of the fraud which is the foundation of the suit, and when the fraud has been concealed, or is of such character as to conceal itself, the statute does not begin to run until the fraud is discovered by, or becomes known to, the party suing or those in privity with him.[209]

Five years later, in *Wood v. Carpenter*,[210] a case addressing an alleged credit fraud, the Court explained that even though the defendant attempted to conceal its fraudulent conduct,

---

[204] *Id.* at 327.
[205] *Id.* at 328.
[206] 670 F. Supp. 527 (S.D.N.Y. 1987), *aff'd on other grounds*, 847 F.2d 1031 (2d Cir. 1988), *cert. denied*, 110 S. Ct. 719 (1990).
[207] 670 F. Supp. at 536. *See also* Clark v. United States, 481 F. Supp. 1086 (S.D.N.Y. 1979), *appeal dismissed*.
[208] 88 U.S. (21 Wall.) 342 (1874).
[209] *Id.* at 349.
[210] 101 U.S. 135 (1879).

> proper diligence could not have failed to find a clue in every case that would have led to evidence not to be resisted. With the strongest motives to action, the plaintiff was supine. If underlying frauds existed, as he alleges, he did nothing to unearth them. It was his duty to make the effort.[211]

These landmark decisions indicate that the Supreme Court has required a potential claimant to exercise reasonable diligence in bringing a cause of action, whatever the effort by the defendant to conceal it.[212] Not to do so would be to defeat the purpose of the limitation period—encouraging prompt claims—in punishing the defendant.[213] Applying this overarching principle to FTCA actions, the period during which a plaintiff must file suit should be calculated to commence from the point when a reasonably diligent person would have recognized a possible cause of action.

B. *Other Conduct Difficult to Discover*

Plaintiffs may have difficulty discovering the "critical facts" of their injuries for reasons other than deliberate concealment by defendants. At what point the claim accrues in such cases is appropriately determined by application of the same diligence discovery rule that is applied in deliberate concealment cases. The medical malpractice context discussed above[214] provides one such example. Application of the rule in this context is straightforward.[215]

Other contexts may pose problems in strictly applying the limitation period similar to those encountered in medical malpractice. For example, plaintiffs wrongfully exposed to toxic substances may not be immediately aware that the exposure will damage their health,[216] or they may not initially be able to iden-

[211] *Id.* at 140.

[212] *See generally* Marcus, *Fraudulent Concealment in Federal Court: Toward a More Disparate Standard?*, 71 GEO. L.J. 829, 875–78 (1983).

[213] *See Kubrick*, 444 U.S. at 117–18; *Wood*, 101 U.S. at 139.

[214] *See supra* notes 76–104 and accompanying text.

[215] *See* Sheehan v. United States, 542 F. Supp. 18, 21 (S.D. Miss. 1982) ("general rule under the Federal Torts Claims Act is that plaintiff's claim accrues when plaintiff's injury manifests itself"; since the plaintiff's injuries manifested themselves more than two years before he presented his FTCA claim, the action was time-barred.)

[216] *See* Allen v. United States, 588 F. Supp. 247 (D. Utah 1984) (exposure to radioactive fallout in Nevada), *rev'd*, 816 F.2d 1417 (1987), *cert. denied*, 484 U.S. 1004 (1988); *In re* "Agent Orange" Prod. Liab. Litig., 506 F. Supp. 737 (E.D.N.Y. 1980) (exposure to "Agent Orange" in Vietnam), *rev'd*, 635 F.2d 987 (1980), *cert. denied*, 454 U.S. 1128 (1981).

tify the exposure as the source of their health problems.[217] Or, apparently harmless activity can later turn out to have been the source of property damage for which plaintiffs seek recovery.[218]

In an extraordinarily difficult case, *Allen v. United States*,[219] the court employed the diligence discovery rule to determine the validity of claims brought by Nevada residents that they had suffered radiation-caused cancer and leukemia due to the government's nuclear testing program. The court found that

> the *Kubrick* standard readily lends itself to a case such as this, in which the injury does not manifest itself until years—sometimes decades—later and in which the critical facts concerning injury or causation are difficult if not impossible to easily ascertain. Construing the statute in this fashion avoids the impossible burden that would be placed on a plaintiff who would otherwise be expected to commence a lawsuit years before he knew he had an injury, knew the source of the injury, and thus knew he had a cause of action.[220]

The court also noted that application of the diligence discovery rule was consistent with the purpose of statutes of limitations, because when an injury takes time to manifest itself, "[g]enuine concern about lost evidence, fading memories, and the passage of time are subordinated to a greater concern that legal wrongs be remedied at the first practical opportunity."[221] Thus, the diligence discovery rule serves as a mechanism for enforcing the objectives of the statute of limitations without precluding claims that cannot fairly be said to have accrued at the time of the injury-causing event.

C. *Continuous Conduct Cases*

The typical FTCA case involves a single injury-causing event, and the accrual analysis in such a case focuses on how long after that event a plaintiff has in which to bring a claim. The fact that the allegedly tortious act may result in repeated or continuous harm will not by itself justify an extension of the

---

[217] *See* Stoleson v. United States, 629 F.2d 1265 (7th Cir. 1980).
[218] *See* Pennback v. United States, 599 F. Supp. 1573 (W.D. Pa. 1985) (action for losses allegedly suffered due to the government's failure to discover construction defects in sewer construction).
[219] 588 F. Supp. 247 (D. Utah 1984).
[220] *Id.* at 341 (citation omitted).
[221] *Id.*

limitation period.[222] The analysis becomes more complicated, however, where the allegedly wrongful conduct itself continues over many years.[223]

If the alleged wrongful acts are separately actionable, the period of limitation is not extended for those acts that accrued more than two years before the action was brought.[224] Allowing a plaintiff to avoid the statute of limitations for such acts would permit the plaintiff to salvage stale claims by characterizing them as elements of a single transaction, and in this way would defeat the very purpose of a statute of repose.[225] In the closely analogous situation of an alleged conspiracy to injure a plaintiff over a period of time, one district court has said:

> In a civil conspiracy action, the conspiracy itself is not actionable, but recovery may be had for the injury caused by specific acts. A person harmed may sue at the time each such act occurs, without having to wait until the termination of the conspiracy. The Statute of Limitations therefore commences to run with respect to each act when it occurs .... Repeated wrongs "are treated as separate rights of action and the Statute of Limitations begins to run as to each ... upon its commission."[226]

---

[222] *See* Ward v. Caulk, 650 F.2d 1144, 1147 (9th Cir. 1981); Maslauskas v. United States, 583 F. Supp. 349, 351 (D. Mass. 1984).

[223] *See generally* Cook v. Pan Am. World Airways, Inc., 771 F.2d 635 (2d Cir. 1985), *cert. denied,* 474 U.S. 1109 (1986); Leonhard v. United States, 633 F.2d 599 (2d Cir. 1980), *cert. denied,* 451 U.S. 908 (1981).

[224] *See generally* Blusal Meats, Inc. v. United States, 638 F. Supp. 824, 829 (S.D.N.Y. 1986) (criminal conspiracy notion, which allows all conspiratorial conduct to be considered as long as the "last overt act" is within the limitation period, held to be inapplicable in civil cases even where conspiracy is pled). Certain specialized statutory causes of action, essentially defining a single wrong of discrimination where a series of discriminatory acts has occurred, or making the existence of a conspiracy independently actionable, may be an exception. In the employment discrimination context, for example, while a plaintiff "may not evade the Title VII stringent time limits merely by characterizing a completed act of discrimination as a 'continuing violation,'" a case-by-case review will be undertaken where a plaintiff asserts that the defendant engaged in a coherent scheme that extended into the limitation period. Drayton v. Veterans Admin., 654 F. Supp. 558, 567 (S.D.N.Y. 1987); *see also* McPartland v. American Broadcasting Cos., Inc., 623 F. Supp. 1334, 1338–39 (S.D.N.Y. 1985). And in a civil RICO claim, there may often be multiple injuries necessitating analysis under "a recognized exception to the general rule" for the accrual of actions on individual injuries, under which "an action is timely as long as the last act evidencing the continuing practice falls within the limitation period." Bankers Trust Co. v. Feldesman, 65 B.R. 470, 490 (S.D.N.Y. 1986).

[225] *See* Singleton v. City of New York, 632 F.2d 185, 192–93 (2d Cir. 1980) (in an action for alleged violation of constitutional rights by police officers, statute of limitations period was not tolled by allegations of conspiracy; the limitation period ran separately for each alleged wrongful act), *cert. denied,* 450 U.S. 920 (1981).

[226] Korry v. International Telephone & Telegraph Corp., 444 F. Supp. 193, 195–96 (S.D.N.Y. 1978) (quoting Baxter v. State, 189 Misc. 525, 72 N.Y.S.2d 337, 340 (Ct. of Claims 1947), *aff'd,* 77 N.Y.S.2d. 796 (1948)). *See also* Maslauskas v. United States,

Where the alleged wrongful conduct does not consist of independently actionable events, and no single event can be identified as the cause of the harm, the conduct may be treated as a "continuous tort" for which the plaintiff has a cause of action that "accrues each day" the wrongful conduct continues.[227] In such cases, the focus is on "when the last tortious act occurred."[228] Although continuous tort cases arise infrequently under the FTCA,[229] they have arisen in the property damage context, in which the damage can be traced to continued exposure to the source of the problem,[230] as well as in the personal injury context, in which the injury results from a series of harmful inflictions.[231] One court has commented that "the *Kubrick* [diligence discovery] rule does not apply" in such cases, noting the difficulty in ascertaining the existence and cause of injury in situations where that injury cannot be ascribed to any particular event: "Since usually no single incident in a continuous chain of tortious activity can 'fairly or realistically be identified as the cause of significant harm,' it seems proper to regard the cumulative effect of the conduct as actionable."[232] The continuous tort doctrine, therefore, stems from a recognition of the difficulties entailed in determining when the injury occurred. The "continous treatment" doctrine is a related doctrine that also alleviates the strict limitation period application in situations in which the tort can be said to occur over an extended period of time.[233] The continuous treatment doctrine results from

583 F. Supp. 349, 351 (D. Mass. 1984) (continued ill effects from prior act do not amount to continuing violation; continued unlawful acts required); *accord Blusal Meats*, 638 F. Supp. at 830 (citing Rutkin v. Reinfeld, 229 F.2d 248, 252 (2d Cir. 1956)); Rodrigues v. Village of Larchmont, New York, 608 F. Supp. 467, 477 n.11 (S.D.N.Y. 1985); Chodos v. FBI, 559 F. Supp. 69, 74 (S.D.N.Y.), *aff'd mem.*, 697 F.2d 289 (2d Cir. 1982).

[227] Rapf v. Suffolk County of New York, 755 F.2d 282, 290 (2d Cir. 1985); Page v. United States, 729 F.2d 818, 823 n.36 (D.C. Cir. 1984).

[228] Gross v. United States, 676 F.2d 295, 300 (8th Cir. 1982).

[229] *See* L. Jayson, Handling Federal Tort Claims: Administrative and Judicial Remedies § 277.04 (1964).

[230] *See* Kennedy v. United States, 643 F. Supp. 1072, 1079 (E.D.N.Y. 1986) (rejecting the government's claim that the limitation period began to run when the plaintiff bought property that was subject to continued erosion caused by the government's construction of stone jetties).

[231] Gross v. United States, 676 F.2d 295 (8th Cir. 1982) (the plaintiff allegedly suffered from emotional distress due to maliciously solicited statements that resulted in a denial of the plaintiff's application to participate in a federal aid program).

[232] Page v. United States, 729 F.2d 818, 821–22 (D.C. Cir. 1984) (quoting Fowkes v. Pennsylvania R.R., 264 F.2d 397, 399 (3d Cir. 1959)).

[233] For an example of the treatment of these issues under state law, see Justice v. Navtig, 238 Va. 178, 381 S.E.2d 8 (1989) (in medical malpractice action against a surgeon, continuing treatment rule applied to eight years of non-negligent treatment that

the courts' recognition of the unfairness of requiring a patient to bring suit while treatment continues.[234] Courts have noted that it would contravene the necessarily trusting nature of the physician-patient relationship to require the patient to interrupt treatment and bring suit against the physician.[235] If the patient felt compelled to investigate possible malpractice to preserve a potential claim, critical treatment might be interrupted.[236] Moreover, the patient who did halt treatment to pursue a claim would face difficulty in obtaining sufficient information to identify the source of the harm, as physicians will naturally be disinclined to reveal all of the important facts about the care they have provided that may have resulted in harm to the patient.[237]

Thus as originally formulated, the "continuous treatment" doctrine was based on the assumption that it was unfair to commence the running of the statute of limitations until the treatment ended. As one court has noted, however, courts "typically assume [the doctrine's] existence and find it inapplicable on the facts."[238] But recent appellate cases have revitalized the doctrine. In *Otto v. National Institute of Health*,[239] the Fourth Circuit applied the doctrine where the plaintiff was undergoing treatment at a new facility that had unique expertise with which to treat her condition. She had virtually no alternative but to exhaust the possibilities of improvement offered by the defendant.[240] In another unusual case, *Ulrich v. Veterans Administration Hosp.*,[241] the plaintiff suffered injuries when he fell or jumped from a smokestack at the defendant's hospital, where he was undergoing treatment for service-related catatonic schizophrenia.[242] The court said that it would be "absurd to

---

followed allegedly negligent operation, and therefore the period of limitations did not begin to run until after the eight-year period of treatment ended).

[234] *See* Ulrich v. Veterans Admin. Hosp., 853 F.2d 1078 (2d Cir. 1988); Otto v. Nat'l Inst. of Health, 815 F.2d 985, 988 (4th Cir. 1987); Tyminksi v. United States, 481 F.2d 257, 264 (3d Cir. 1973); Kelly v. United States, 554 F. Supp. 1001, 1003 (E.D.N.Y. 1983).

[235] *Ulrich*, 853 F.2d at 1080; *Brown*, 353 F.2d 578, 580 (9th Cir. 1965); Kossick v. United States, 330 F.2d 933 (2d Cir. 1964), *cert. denied*, 379 U.S. 837 (1964).

[236] *Ulrich*, 853 F.2d at 1080; *Tyminski*, 481 F.2d at 264.

[237] Ashley v. United States, 413 F.2d 490, 493 (9th Cir. 1969).

[238] Kelly v. United States, 554 F. Supp. 1001, 1003 (E.D.N.Y. 1983); *see, e.g., Tyminksi*, 481 F.2d at 264 n.5; Dundon v. United States, 559 F. Supp. 469, 472–73 (E.D.N.Y. 1983); DeGirolamo v. United States, 518 F. Supp. 778, 781 (E.D.N.Y. 1981); Mortensen v. United States, 509 F. Supp. 23, 29–30 (S.D.N.Y. 1980).

[239] 815 F.2d 985 (4th Cir. 1987).

[240] *Id.* at 988–89.

[241] 853 F.2d 1078 (2d Cir. 1988).

[242] *Id.* at 1079.

require [the plaintiff] to interrupt [the defendant's] corrective treatment in order to commence legal proceedings."[243]

In *Otto* and *Ulrich*, requiring the plaintiff to bring suit while still undergoing treatment was obviously inappropriate. The appropriate result was reached through the continuous treatment doctrine; but the same result would be achieved by applying the diligence discovery rule. The concerns that led to the enunciation of the "continuous treatment" doctrine—*i.e.*, the nature of the physician-patient relationship, the desirability of continuing apparently necessary treatment, the difficulty in gathering important facts—would all be taken into account in the diligence discovery analysis regarding what constitutes reasonable inquiry. In the appropriate case, that analysis will result in a delay in the accrual of the cause of action. Thus the continuous treatment doctrine should be viewed not as an exception to the due diligence requirement, but rather "a factor in determining whether that requirement has been met."[244]

In *Wehrman v. United States*,[245] the court considered applying a *Kubrick* diligence discovery rule in the continuous treatment context. In *Wehrman*, the lower court had ruled that the limitation period barred a plaintiff from bringing a suit based on treatment provided to him over the course of twenty-two years in a veterans hospital, during which his condition progressively worsened. According to the court of appeals, the district court had concluded that the plaintiff "failed to exercise reasonable diligence in becoming aware of a possible claim."[246] The government argued that the continuous treatment doctrine was "no longer viable following the Supreme Court's decision in *Kubrick* . . . [, which] shifted the focus to the exercise of reasonable diligence by the plaintiff in discovering the injury."[247]

Although the court of appeals did not apply a diligence inquiry in *Wehrman*, it joined with the lower court in rejecting the "suggestion that the continuing treatment doctrine had been completely eviscerated by *Kubrick*."[248] The court relied heavily on continuous tort, rather than continuous treatment, cases in its analysis,[249] and recited the continuous tort principle that

[243] *Id.* at 1081.
[244] Kelly v. United States, 554 F. Supp. 1001, 1004 (E.D.N.Y. 1983).
[245] 830 F.2d 1480 (8th Cir. 1987).
[246] *Id.* at 1486.
[247] *Id.*
[248] *Id.*
[249] *Id.* at 1483–86.

"[w]here the tortious conduct is of a continuing nature, the *Kubrick* rule does not apply."[250] Yet it also noted certain other elements, such as deliberate concealment, which factor into a diligence determination:

> We . . . reject the district court's conclusion that Wehrman's claim is barred because he failed to exercise reasonable diligence in becoming aware of a possible claim. We add in passing that even if a diligence inquiry would apply in a continuing treatment context, the VA's affirmative actions to dissuade Wehrman from surgery ought at least to be a factor, along with his knowledge of the deterioration and severity of his condition, in analyzing whether Wehrman knew or should have known that there may have been negligence.[251]

Thus the court suggests that its ruling may have been the same in *Wehrman* if it had applied the diligence discovery rule. The *Wehrman* court appeared to be struggling to avoid the harsh consequences of a strict application of the statute of limitations, which the court suggested must follow unless the continuous treatment doctrine was applied. But, as explained above, in an appropriate case the diligence discovery rule takes into account the same concerns that lead to application of the continuous treatment doctrine, as well as additional concerns that favor the plaintiff and serve the interests of fundamental fairness.

IV. Amending the FTCA to Reflect the Diligence Discovery Standard

Courts have been generally reluctant to employ diligence discovery principles to resolve statute of limitations problems under the FTCA beyond the malpractice context. The common law has arrived at currently accepted applications of the diligence discovery test noncomprehensively. While some commentators have read *Kubrick* as stating a universal federal limitations accrual principle,[252] accrual questions in many FTCA cases are still decided on other bases.[253] To promote uniformity and fairness in the application of the doctrine, and to serve the

[250] *Id.* at 1486 (quoting Gross v. United States, 676 F.2d 295, 300 (8th Cir. 1982)).
[251] *Id.* at 1486.
[252] *E.g.*, S. Nahmod, Civil Rights and Civil Liberties Litigation § 4.15 (1986). *See also* Sowers v. Bradford School Dist., 694 F. Supp. 125, 136 (W.D. Pa. 1988).
[253] *See supra* notes 166–173 and accompanying text.

remedial purpose of the FTCA, Congress should consider an amendment to the limitation provision that expressly authorizes the use of diligence discovery considerations under certain circumstances.

The need for greater equity and uniformity is sufficient reason for amending section 2401(b). But such amendment also may enable Congress to economize its own resources and more fully effectuate two original purposes of the Act: (1) relief of the congressional burden of resolving tort claims against the government through private legislation;[254] and (2) establishment of a remedy against the government as a matter of right rather than of sovereign grace. Section 2401(b) has hindered in part the fulfillment of each of these objectives.

First, although private claims that may be brought under the Act are banned from introduction into Congress,[255] the spate of exceptions and limitations to its coverage allow a range of claims to continue to be brought as private bills, which are not subject to a two-year statute of limitations.[256] Moreover, plaintiffs whose claims are time-barred may still introduce private tort claim bills in order to waive the expired statute of limitations.[257] Thus plaintiffs which would face a strict application of the limitation period in court will instead pursue private bills; as a result, many tort claim bills may be introduced into each congressional session.

Private bills must follow the same general enactment procedures as public bills. Specifically, the claimant must find a spon-

---

[254] *See* S. REP. No. 1400, 79th Cong., 2d Sess. 2, 7 (1946).

[255] 60 Stat. 831 (1946).

[256] Procedural rules of the House Committee on the Judiciary include a limitations provision governing the introduction of private claims bills: unless waived by a vote of two thirds of the appropriate subcommittee, private bills must be introduced within 15 years of the date the claim first accrued. HOUSE COMM. ON THE JUDICIARY, 101ST CONG., SUPPLEMENTAL RULES OF PROCEDURE FOR PRIVATE CLAIMS BILLS 4–5 (Comm. Print 1989). *See also supra* text accompanying notes 29–34.

[257] *See* Note, *Private Bills in Congress*, 79 HARV. L. REV. 1684, 1695–96 (1966) (citing examples of waiver of an expired statute of limitations by private bills). William Kubrick, for instance, who lost his $320,000 judgment when the Supreme Court ruled that his claim was time-barred, was the subject of five private bills introduced in four consecutive years. *See* S. 232, 98th Cong., 1st Sess.; S. 1619, 97th Cong., 1st Sess.; H.R. 4836, 97th Cong., 1st Sess.; H.R. 7151, 96th Cong., 2d Sess.; S. 2169, 96th Cong., 1st Sess.. None of these bills, however, made it out of the House or Senate Committees on the Judiciary.

In the past five years only two private bills based on § 2401(b) have been passed. *See* Priv. L. No. 99-15, 100 Stat. 4319 (1986); Priv. L. No. 99-18, 100 Stat. 4320 (1986). Neither of these resolved the substantive claim involved, but rather they waived § 2401(b) by vesting the district courts with jurisdiction to hear the claims. *Id.*

sor who will draft and introduce the legislation.[258] The bill is then referred to the appropriate committee, where staff begin an investigation into the claim's merits.[259] If the bill is reported favorably,[260] it will generally pass through that house. Due to the vagaries of the system, however, only about fifteen to twenty percent of private bills are ever enacted,[261] and currently most of those are immigration-related.[262] Even though few private bills based on tort claims otherwise barred by section 2401(b) are enacted,[263] all of those that are introduced unnecessarily command our lawmakers' time and attention.

An amendment that broadens the scope of the FTCA would reduce the onus of private bills.[264] Codifying a diligence discovery test will allow many claims under the FTCA that might otherwise be time-barred to proceed through the court system rather than the legislature, thus effectuating one of the original goals of the FTCA while conserving congressional resources. Such an amendment will also serve a second congressional purpose behind the Act—establishing a tort remedy against the United States as a matter of right rather than grace—by clarifying the rights of FTCA plaintiffs to proceed in the courts under the limitation provision and thus keeping many claims out of the discretionary private bill process.

A. *A Proposed Core Amendment*

Many states have adopted a common law diligence discovery test, similar to that of the federal courts which led to *Kubrick*,

---

[258] *See* G. GALLOWAY, THE LEGISLATIVE PROCESS IN CONGRESS 529–35 (1955).

[259] Private bills can be introduced into either house, although roughly 85% are introduced in the House of Representatives. *Id.* at 533. Once introduced, most domestic legislation is referred to the Committee on the Judiciary of the appropriate house. *See* Note, *Private Bills in Congress*, *supra* note 257, at 1688, 1691. Separate standing subcommittees handle immigration-related bills and claims against the United States. *Id.*

[260] Bills that are reported favorably generally have the support of the executive department or agency that would otherwise have jurisdiction over the claim. *See* Gellhorn & Lauer, *Congressional Settlement of Tort Claims Against the United States*, 55 COLUM. L. REV. 1, 9–13 (1955).

[261] *Id.* at 2.

[262] *See* HOUSE JOURNAL, 98th Cong., 2d Sess. 2321 (1984); HOUSE JOURNAL, 98th Cong., 1st Sess. 1971 (1983); 98 Stat. 3417–38 (1984); 97 Stat. 1483–86 (1983).

[263] *See supra* note 257.

[264] This of course holds true for any other exception to the availability of an FTCA cause of action.

to determine when causes of action accrue.[265] However, some sixteen states have enacted statutory provisions setting forth versions of this test for one or more categories of causes of action.[266]

Congress itself has shown a willingness to consider further amendments to the FTCA limitations provisions.[267] Hence it is appropriate to sketch versions of a statutory proposal that could clarify and implement the diligence discovery standard for accrual of tort claims against the United States. Presented below are three alternative amendments to § 2401(b), which reserve to a varying extent certain accrual issues for common law treatment (changes to existing language and additions are noted in bold print):

(b) **Except as provided in subsections (c) and (d) of this section, a** tort claim against the United States shall be forever barred unless it is presented in writing to the appropriate federal agency within two years after such claim accrues or unless action is begun within six months after the date of mailing, by certified or registered mail, of notice of final denial of the claim by the agency to which it was presented. **A tort claim against the United States shall be deemed to have accrued** [insert Alternative A, B, or C]

Alternative A:

when the injury is first discovered or in the exercise of reasonable care [diligence] **should have been discovered.**

---

[265] *See generally* cases collected in Annotation, *Medical Malpractice: Statute of Limitations in Wrongful Death Action Based on Medical Malpractice*, 70 A.L.R. 4th 535 (1989); Annotation, *Time of Discovery as Affecting Running of Statute of Limitations in Wrongful Death Action*, 49 A.L.R. 4th 972 (1986); Annotation, *Limitation of Actions: Time of Discovery of Defamation as Determining Accrual of Action*, 35 A.L.R. 4th 1002 (1985).

[266] *See* CAL. CIV. PROC. CODE §§ 340.2, 340.5 (West 1982); CONN. GEN. STAT. ANN. § 52-584 (West Supp. 1990); DEL. CODE ANN. tit. 18, § 6856 (1989); FLA. STAT. ANN. § 95.11(4) (West 1982 & Supp. 1990); HAW. REV. STAT. § 657-7.3 (1988); IDAHO CODE § 5-219 (1990); IOWA CODE ANN. § 614.1(9) (West Supp. 1990); KAN. STAT. ANN. §§ 60-513, 60-513b (Supp. 1990); KY. REV. STAT. ANN. § 413.140(2) (Michie/Bobbs-Merrill 1988); NEB. REV. STAT. § 25-224(1), (5) (1985); N.H. REV. STAT. ANN. § 508:4 (1988); N.C. GEN. STAT. § 1-52(16) (1983); OR. REV. STAT. § 12.110(4) (1989)); OR. REV. STAT. § 12.115 (1989); S.C. CODE ANN. § 15-3-535 (Law. Co-op. Supp. 1989); S.C. CODE ANN. § 15-3-545 (Law. Co-op. Supp. 1989); VT. STAT. ANN. tit. 12, § 521 (Supp. 1989); WIS. STAT. ANN. § 893.55(1) (West 1983).

[267] *See supra* notes 30–43 and *infra* notes 290–292.

This amendment adopts the *Kubrick* standard,[268] and its language primarily draws upon that of Oregon's statute governing accrual of medical malpractice actions.[269] Many other state provisions use similar wording.[270]

Although the language of this amendment is easy to understand and sets out the diligence discovery rule, this straightforward and simple version is subject to overly narrow interpretation because it does not distinguish between knowledge of the injury itself and knowledge of its cause in fact. In *Kubrick* the Court stressed that knowledge of both an injury *and* its cause is required before a claim accrues under the FTCA.[271] Thus, while this alternative appears sufficient to deal with cases of hidden or latent injury, its failure to deal specifically with the two elements of knowledge that are generally taken to trigger claim accrual may generate confusion when the courts apply the statute, particularly in those medical malpractice cases where the injury may have been apparent for years before the plaintiff reasonably could become aware of the cause of that injury.[272]

Further, the language of this amendment draft is not sufficiently explicit in addressing cases in which an injury does not become apparent until years after the initial act or incident that caused the harm occurred,[273] or cases in which an injury or its cause are concealed from the plaintiff.[274] Moreover, under any

---

[268] 444 U.S. at 122.

[269] OR. REV. STAT. § 12.110(4) (1989).

[270] *See* CAL. CIV. PROC. CODE § 340.5 (West 1982); CONN. GEN. STAT. ANN. § 52-584 (West Supp. 1990); FLA. STAT. ANN. § 95.11(4) (West 1982 & Supp. 1990); HAW. REV. STAT. § 657-7.3 (1988); KY. REV. STAT. ANN. § 413.140(2) (Michie/Bobbs-Merrill Supp. 1990); S.C. CODE ANN. § 15-3-545 (Law. Co-op. Supp. 1989); WIS. STAT. ANN. § 893.55(1) (West 1983).

[271] 444 U.S. at 125.

[272] *See, e.g.*, Barnhart v. United States, 884 F.2d 295 (7th Cir. 1989) (patient learned years later that his tardive dyskinesia was caused by administration of neuroleptic drugs), *cert. denied*, 110 S. Ct. 2561 (1990); Oberlin v. United States, 727 F. Supp. 946 (E.D. Pa. 1989) (issue of material fact existed as to when parents learned that minor child's cerebral palsy could be result of physician's failure to treat premature rupture of membranes appropriately); Young v. United States, 698 F. Supp. 393 (D. Mass. 1988) (patient did not have reasonable opportunity to discover cause of bone infection following knee surgery until subsequent examination by orthopedic surgeon); Nemmers v. United States, 681 F. Supp. 567 (C.D. Ill. 1988) (parents learned that minor child's cerebral palsy could have been caused by negligence of physician years after child's birth), *aff'd*, 870 F.2d 426 (7th Cir. 1989).

[273] *See, e.g.*, Barrett v. United States, 689 F.2d 324 (2d Cir. 1982), *cert. denied*, 462 U.S. 1131 (1983); Stoleson v. United States, 629 F.2d 1265 (7th Cir. 1980); Allen v. United States, 588 F. Supp. 247 (D. Utah 1984), *rev'd*, 816 F.2d 1417 (10th Cir. 1987); *In re* "Agent Orange" Prod. Liab. Litig., 506 F. Supp. 757 (E.D.N.Y. 1980).

[274] *See supra* text accompanying notes 177–213; *see also* McDonald v. United States, 843 F.2d 247 (6th Cir. 1988); Cogburn v. United States, 717 F. Supp. 958 (D. Mass. 1989); Santana v. United States, 693 F. Supp. 1309 (D.P.R. 1988).

diligence discovery system for determining the accrual of causes of action, the length of the limitation period may be very short or potentially infinite, since the accrual date in such actions would depend solely upon the court's determination of what constitutes due or reasonable diligence by the plaintiff. The language of Alternative A provides little guidance to the courts in these situations, thereby increasing the possibility of inconsistent adjudication.

Given that federal jurisprudence in FTCA and analogous cases has emphasized the specific elements of knowledge a plaintiff may have, any amendment to 28 U.S.C. section 2401(b) should deal explicitly with the elements of knowledge required. An elaborated statutory formula will permit more effective implementation of the concepts articulated by the Court in *Kubrick* in determining when a cause of action accrues.

Alternative B:

when the claimant discovers or in the exercise of due [reasonable] **diligence should have discovered both the existence of the injury giving rise to the cause of action** [claim] **and the cause** [in fact] **of that injury.**

This form of amendment explicitly states that a cause of action under the Federal Tort Claims Act accrues when a plaintiff has actual or constructive knowledge of the injury and its cause—a more complete statement of the diligence discovery rule articulated by the *Kubrick* Court.[275] The amendment employs an objective standard for determining whether a plaintiff has exercised due diligence to learn whether a cause of action exists, thus following the Court's suggestion in *Kubrick* that when both the existence of an injury and its cause are known, a plaintiff has a duty to seek advice in the medical and legal community.[276]

The language of this amendment primarily draws upon Connecticut and Florida statutes governing accrual of actions for medical malpractice,[277] borrowing the phrase "due diligence" from the Court's language in *Kubrick*.[278] The phrase "reasonable

[275] 444 U.S. at 122.
[276] *Id.* at 123–24.
[277] *See* CONN. GEN. STAT. ANN. § 52-584 (West Supp. 1990); FLA. STAT. ANN. § 95.11(4) (West 1982 & Supp. 1990).
[278] 444 U.S. at 116.

diligence" could be used as well, to accord with the diligence discovery rule's articulation elsewhere in *Kubrick* and in several state statutes.[279]

The word "cause" is used in this amendment in the *Kubrick* sense,[280] to indicate that knowledge of the probable, rather than the "legal," cause of the injury is necessary for a cause of action to accrue under the FTCA. The words "cause in fact" may be preferable, however, to insure that the amendment is not interpreted to require (as do the laws of several states) knowledge or evidence of negligent or wrongful conduct by another in order to trigger running of the statutory period. Several jurisdictions defer running the statutory period until there is awareness of the legal cause. That awareness is variously described as knowledge by plaintiff of wrongdoing, negligence, actionability, or all the elements of a cause of action.[281] However, this approach is unduly protective of plaintiffs. A person who is aware of the fact of injury and its factual cause should instead be required to make reasonable investigation for self-protection and need not be shielded until demonstration of actionabililty fortuitously occurs.

While the language in Alternative B specifies these two elements as constituting knowledge sufficient for a cause of action to accrue under the FTCA, this draft provision stops short of spelling out the rule for situations in which an injury or its cause cannot be reasonably discovered until some time after the initial act giving rise to the action. Problems may arise in cases where either the injury itself is not discovered until years after the initial act or incident, or the injury or its cause are concealed from the plaintiff.[282] The amendment is similarly silent on the question of a plaintiff's rights and responsibilities in continuous

---

[279] *See, e.g., id.* at 120–24; CAL. CIV. PROC. CODE §§ 340.2, 340.5 (West 1982); DEL. CODE ANN. tit. 18, § 6856 (1989); HAW. REV. STAT. § 657-7.3 (1988); IOWA CODE ANN. § 614.1(9) (West Supp. 1990); N. H. REV. STAT. ANN. § 508:4 (Supp. 1988); and WIS. STAT. ANN. § 893.55(1) (West 1983).

[280] 444 U.S. at 122.

[281] *See* Bussineau v. President and Directors of Georgetown College, 518 A.2d 423, 435 (D.C. App. 1986) (declining to follow *Kubrick* discovery rule in a dental malpractice action, holding that cause of action accrued when the plaintiff knew or should have known by reasonable diligence of the injury, of its cause in fact, and of some evidence of wrongdoing). The court in *Bussineau* noted that North Dakota and Hawaii have rejected the diligence discovery rule as set out in *Kubrick. Id.* at 431–32 (citing Anderson v. Shook, 333 N.W. 2d 708 (N.D. 1983) (requiring knowledge of the injury, its cause, and of "'possible negligence'" before a cause of action accrues); Jacoby v. Kaiser Found. Hosp., 1 Haw. App. 519, 622 P.2d 613 (1981) (requiring knowledge of damage, of violation of a duty, and of a causal connection between the violation of the duty and the damage before the cause of action accrues)).

[282] *See* the cases cited *supra* notes 273–274.

tort cases. Thus this approach to drafting the amendment does not offer much more guidance to courts than Alternative A and likewise may preserve the possibility of inconsistent adjudication in these situations.

Alternative C:

(1) when the act [or omission] **giving rise to the cause of action first causes** [substantial] **injury, or (2) if the fact of injury and its cause** [in fact] **are not reasonably discoverable** [ascertainable] **until some time after the initial act** [or omission]**, when the claimant** [injured party] **discovers** [ascertains] **or in the exercise of due** [reasonable] **diligence should have discovered** [ascertained] **the fact of injury and its cause** [in fact].

This final variant incorporates the discovery rule stated in *Kubrick* and includes a provision for situations in which it may be difficult for a plaintiff exercising reasonable diligence to discover the cause of action. The amendment sets out the two-part test employed by the Court in *Kubrick* to determine accrual of a cause of action, and employs an objective standard.[283] It also provides for continuous torts by accruing claims at the time of initial injury, unless circumstances prevent reasonable discovery.

This version draws not upon a statute governing accrual of medical malpractice actions but instead upon a Kansas statute governing accrual of various personal injury actions.[284] It is intended that the statute will provide guidance to courts in many areas beyond medical malpractice. The Kansas statute uses the word "ascertain" rather than "discover."[285] However, the word "discover" conforms more closely with the language in *Kubrick*,[286] without any apparent loss of clarity. Prior discussions of the use of the words "reasonable diligence" over "due diligence" and "cause in fact" over "cause" would apply to this amendment as well.

The Kansas statute uses the phrase "substantial injury" in its provision governing accrual of several types of actions, including trespass on real property; taking, detaining, or injuring personal property; fraud; injury to the rights of another not arising

---

[283] *See* 444 U.S. at 122–24.
[284] KAN. STAT. ANN. § 60-513 (Supp. 1990).
[285] *Id.*
[286] 444 U.S. at 120, 122.

on contract; and wrongful death. However, the statute uses only the word "injury" in its provision governing accrual of medical malpractice actions.[287] The Supreme Court of Kansas has construed the phrase "substantial injury" to mean "actionable injury," stating that to trigger the statute of limitations,

> the term "substantial injury" in the statute does not require an injured party to have knowledge of the full extent of the injury . . . . Rather, it means the victim must have sufficient ascertainable injury to justify an action for recovery of the damages, regardless of extent. An unsubstantial injury as contrasted to a substantial injury is only a difference in degree, i.e., the amount of damages. That is not a legal distinction.[288]

Characterizing an injury as substantial strengthens the notion that knowledge of the injury is enough to require the victim to investigate the circumstances and ascertain whether the injury is actionable.

The phrase "act or omission" is found in the statutes of several states governing accrual of various personal injury actions, particularly medical malpractice actions.[289] This formulation captures the appropriate range of conduct that may lead to a cause of action under the FTCA.

Alternative C, while somewhat cumbersome, expressly addresses situations in which an injury or its cause is difficult to discover, even in the exercise of reasonable diligence. Courts could apply this amendment to deal effectively with situations in which an injury is not discovered until long after the initial act or incident giving rise to the action, or in which an injury or its cause is concealed from a plaintiff, and thus preserve fundamental fairness without introducing excessive uncertainty into the waiver of immunity.

B. *Dealing with Other "Harsh Results": Proposed Amendments Before Congress*[290]

Under certain circumstances, the administration of limitation provisions can cause harsh or unjust results. Congress is cur-

---

[287] *See* KAN. STAT. ANN. §§ 60-513(b), (c) (Supp. 1990).
[288] Roe v. Diefendorf, 236 Kan. 218, 222, 689 P.2d 855, 859 (1984).
[289] *See, e.g.*, HAW. REV. STAT. § 657-7.3 (1988); IDAHO CODE § 5-219 (1990); N.H. REV. STAT. ANN. § 508:4 (1983 & Supp. 1989); S.C. CODE ANN. § 15-3-545 (Law Co-op. Supp. 1989).
[290] *See* Barren v. United States, 839 F.2d 987, 997 (3d Cir. 1988) (Becker, J., dissenting).

rently considering two amendments to the limitation provisions of the FTCA that deal specifically with important and repetitive problems of this nature. These pending bills would toll the FTCA statute of limitations for minors and for persons under legal disability. We would include them as subsections (c) and (d) of our amendment proposed above, as follows:

(c) A tort claim against the United States of any person who is under the age of 18 years at the time the claim accrues may be presented to the appropriate federal agency not later than two years after such person reaches the age of 18 years.[291]

(d) A tort claim against the United States of any person who is under legal disability at the time the claim accrues may be presented to the appropriate federal agency not later than two years after the disability ceases.[292]

Many states have created exceptions that toll statutes of limitation in other specific circumstances, and over time Congress may elect to add further specific provisions if the volume and problematic nature of any category of case appear to warrant such amendment. Exceptions found in state statutes include medical malpractice actions arising out of placement of a foreign object that has no therapeutic purpose or effect in a patient's body,[293] exposure to phenoxy herbicides,[294] exposure to asbestos,[295] ionizing radiation injury,[296] legal malpractice,[297] nuclear incidents involving the release of radioactive material,[298] and certain forms of securities transactions.[299] Congress may wish to consider whether to amend section 2401(b) to include exceptions in any of these circumstances. Congress might also codify

---

[291] H.R. 3260, 101st Cong., 1st Sess., 135 CONG. REC. H5674-09 (1989).
[292] H.R. 3261, 101st Cong., 1st Sess., 135 CONG. REC. H5674-09 (1989).
[293] *See, e.g.*, CAL. CIV. PROC. CODE § 340.5 (West 1982); IDAHO CODE § 5-219 (1990); IOWA CODE ANN. § 614.1(9) (West Supp. 1990); S.C. CODE ANN. § 15-3-545 (Law. Co-op. Supp. 1989); VT. STAT. ANN. tit. 12, § 521 (Supp. 1989); WIS. STAT. ANN. § 893.55(3) (West 1983); for actions arising out of a latent defect, see FLA. STAT. ANN. § 95.11(3)(c) (West 1982).
[294] *See* FLA. STAT. ANN. § 95.11(4)(f) (West Supp. 1990).
[295] *See, e.g.*, CAL. CIV. PROC. CODE § 340.2 (West 1982); NEB. REV. STAT. § 25-224(5) (1985).
[296] *See* KAN. STAT. ANN. § 60-513b (Supp. 1990).
[297] *See* CAL. CIV. PROC. CODE § 340.6 (West 1982).
[298] *See* OR. REV. STAT. § 12.110(5) (1989).
[299] *See* FLA. STAT. ANN. § 95.11(4)(e) (West Supp. 1990).

the federal common law exception for fraudulent concealment,[300] which is frequently found in state statutes as well.[301]

But if the courts appropriately apply a diligence discovery rule to all actions filed under the FTCA, most statutory exceptions to the two-year limitation period would be unnecessary. Under this rule, a cause of action does not accrue until a plaintiff is aware of both the injury and its cause in fact.[302] Thus, in situations where a person was exposed to a toxic chemical, for example, the cause of action would not accrue until the plaintiff discovered or should have discovered the injury and its cause, though several years may pass before either of these becomes evident or reasonably discoverable. Thus, while we do not argue against the inclusion of specific provisions for repetitive situations, we simply note that a diligence discovery rule can effectively deal with many situations that have produced "harsh results" in the past.

C. *Resolving the Government Status Problem*

A potential problem for plaintiffs under section 2401(b) is that the statute does not require that plaintiffs be aware of the government-actor status of defendants before their claims accrue under the FTCA. Problems may arise when a plaintiff is injured, knows of that injury and its cause in fact, but is unaware that the person who inflicted that injury is a government employee acting within the scope of employment at the time of the incident, making the United States the proper defendant in a lawsuit and subjecting the action to the requirements of the FTCA. A plaintiff's claim may be time-barred if the administrative claim is not filed within the two-year period following accrual as required by the FTCA,[303] even though the plaintiff was unaware that the defendant was a government employee. This problem

---

[300] *See* Barrett v. United States, 689 F.2d 324, 327 (2d Cir. 1982) (quoting Fitzgerald v. Seamans, 553 F.2d 220, 228 (D.C. Cir. 1977)).

[301] *See, e.g.*, CONN. GEN. STAT. ANN. § 52-595 (West 1960); IDAHO CODE § 5-219 (1990); VT. STAT. ANN. tit. 12, § 521 (Supp. 1988) (for fraudulent and/or knowing concealment); CAL. CIV. PROC. CODE § 340.5 (West 1982); FLA. STAT. ANN. § 95.11(4)(b) (West 1982) (for fraud, (intentional) concealment, or intentional misrepresentation of fact); WIS. STAT. ANN. § 893.55(2) (West 1983) (for concealment in medical malpractice actions); OR. REV. STAT. § 12.110(4) (1989) (for fraud, deceit, or misleading representation in medical malpractice actions).

[302] *See Kubrick*, 444 U.S. at 122.

[303] 28 U.S.C. § 2401(b) (1988).

has most frequently arisen in actions for injuries sustained in automobile accidents,[304] but it has also arisen in medical malpractice.[305]

The circuits are divided over whether a court has jurisdiction over an FTCA suit in which the plaintiff did not file a claim with the appropriate federal agency within two years of the injury because the defendant's government status was not known. In the context of automobile accidents, the majority of courts have dismissed such actions, despite the plaintiff's lack of knowledge.[306]

Courts that have dismissed actions for failure to file a claim with the appropriate federal agency within the time limit have typically reasoned that a reasonably diligent plaintiff would be expected to discover the defendant's government-actor status.[307] In *Bradley v. United States*, for example, the Third Circuit noted that it "[does] not decide whether there might be a basis for some relaxation [of the filing requirements of the Federal Tort Claims Act] if it were in fact impossible for a diligent claimant to present a notice within two years of the claim accruing."[308] The court found, however, that

> [t]his was a routine automobile accident case in which the government employee was immediately identified, though his status was not. Thus, if [the plaintiffs] had promptly filed their action there is no doubt that with minimal discovery

---

[304] *See, e.g.*, Bradley v. United States, 856 F.2d 575 (3d Cir. 1988); Henderson v. United States, 785 F.2d 121 (4th Cir. 1986); Wilkinson v. United States, 677 F.2d 998 (4th Cir. 1982), *cert. denied*, 459 U.S. 906 (1982); Wollman v. Gross, 637 F.2d 544 (8th Cir. 1980); Kelley v. United States, 568 F.2d 259 (2d Cir. 1978), *cert. denied*, 439 U.S. 830 (1978).

[305] *See, e.g.*, Gould v. United States Dep't of Health & Human Serv., 884 F.2d 785 (4th Cir. 1989); Flickinger v. United States, 523 F. Supp. 1372 (W.D. Pa. 1981).

[306] *See, e.g.*, *Bradley*, 856 F.2d 575; Houston v. United States Postal Serv., 823 F.2d 896 (5th Cir. 1987); *Henderson*, 785 F.2d 121; *Wilkinson*, 677 F.2d 998; *Wollman v. Gross*, 637 F.2d 544. *But see* Staple v. United States, 740 F.2d 766 (9th Cir. 1984); *Kelley*, 568 F.2d 259.

[307] *Id.*; *see Houston*, 823 F.2d at 902 (stating that a plaintiff who knows or should know that a driver was a government employee must exhaust the FTCA's administrative requirements and then commence suit against the government on time); *see also Henderson*, 785 F.2d 121 (statute was not tolled because the plaintiff's counsel was informed by U.S. Attorney of need to file administrative claim before proceeding in U.S. District Court); *Wilkinson*, 677 F.2d 998 (statutory period was not tolled for a plaintiff who possessed sufficient knowledge to put him on inquiry as to whether the defendant, a Navy boatswain on active duty, was operating a motor vehicle within the scope of his employment); *Wollman v. Gross*, 637 F.2d 544 (statutory period was not tolled because the plaintiff was aware that the defendant was employed by government agency at the time of the accident and could have been reasonably expected, with assistance of legal counsel, to research the defendant's status for consideration of FTCA claim).

[308] *Bradley*, 856 F.2d at 579.

> they could have ascertained [the defendant] was a government employee in the scope of his employment so that timely notice could have been given under the Federal Tort Claims Act. Instead they chose to wait. While they were free to do so, they delayed at their own peril . . . . Thus while our result may seem harsh we are compelled to reach it.[309]

Those courts that have allowed plaintiffs to proceed despite noncompliance with the FTCA's administrative claim requirement (notably the Second Circuit) have in effect applied a diligence discovery rule, concluding that in these cases, the plaintiffs did not and could not have known that the defendants were government employees acting within the scope of their employment. In *Kelley v. United States*,[310] a case decided before *Kubrick*, the Second Circuit held that an action originally filed against a federal employee in state court and subsequently removed to federal court was not barred, although the plaintiffs had failed to present an administrative claim within the period prescribed by the FTCA.[311] The court stated that:

> In exactly the most excusable and understandable case—the case of the plaintiff who sues in ignorance of the fact that the defendant was a federal driver operating within the scope of his employment—requiring an administrative filing produces the most unjust refinement of interpretation: the plaintiff must have filed a claim that he did not know he had; his suit must be dismissed unless plaintiff can prove that the Government was wrong in certifying that the federal employee was acting within the scope of his employment.[312]

The court in *Kelley* also was concerned that the government might "lull plaintiffs into a false sense of security by waiting until plaintiffs' time to file an administrative claim had expired and thereupon move to be substituted [as a party defendant] and to dismiss,"[313] and stated that

> [t]he statute can not be thought to contemplate that the defense of the case by the United States will consist in moving to dismiss it because no administrative claim . . . was filed. That will typically have been the case; few, if any, plaintiffs will have sued the federal driver knowing that the

---

[309] *Id.*
[310] 568 F.2d 259 (2d Cir. 1978).
[311] *Id.*
[312] *Id.* at 266.
[313] *Id.* at 262.

> suit might properly have been commenced against the United States.[314]

While several courts have declined to follow *Kelley*,[315] two courts have subsequently held that failure to file an administrative claim within the limitation period did not bar a plaintiff's claim because the plaintiff had no knowledge or reason to suspect that the defendant was a government employee acting within the scope of employment.[316] One district court expressly formulated a rule to deal with these situations:

> [I]n a case in which the plaintiff prior to filing suit knew or had reason to know that the driver was (1) a federal employee (2) acting within the scope of his employment at the time of the accident, the requirement of Section 2675 [that an administrative claim be filed] applies. The plaintiff is required to seek administrative remedies; filing in state court is not a means of avoiding this requirement. Where the driver of a motor vehicle is sued individually in state court because the plaintiff did not know and had no reason to know that the defendant was (1) a federal employee (2) on federal business at the time of the accident and the United States subsequently removes the action to federal court under Section 2679, no exhaustion of administrative remedies is required.[317]

The court said that "[t]his formulation of the rule fits the congressional purpose more closely than the Second Circuit's broader rule [as stated in *Kelley*]."[318]

In one medical malpractice case in which a plaintiff failed to file an administrative claim within the time limit, maintaining that she did not know the defendant was a government em-

---

[314] *Id.* at 265. *Kelley* was motivated by the court's concern for the innocent plaintiff who had no knowledge or reason to suspect that the defendant is a government employee until it was too late, but the facts of the case do not appear to support this premise. There the plaintiffs may have been on notice that the defendant was a government employee. Two months after the accident, an agent of the Department of Agriculture charged with investigating the accident interviewed the plaintiffs, and 16 months after the accident, the defendant testified at his deposition that he had been working the day of the accident, and the fact of his employment with the government was explicitly noted on the record. *Id.* at 261.

[315] *See, e.g.*, Bradley v. United States, 856 F.2d 575 (3d Cir. 1988); Houston v. United States Postal Serv., 823 F.2d 896 (5th Cir. 1987); Kozel v. Dunne, 678 F. Supp. 450 (D.N.J. 1988); Gonzales v. United States Postal Serv., 543 F. Supp. 838 (N.D. Cal. 1982); Harris v. Burris Chem., Inc., 490 F. Supp. 968 (N.D. Ga. 1980); LePatourel v. United States, 463 F. Supp. 264 (D. Neb. 1978).

[316] *See* Van Lieu v. United States, 542 F. Supp. 862 (N.D.N.Y. 1982) (military status of driver was in no way communicated to the plaintiff until time following expiration of two-year period of limitation for filing administrative claim); *Harris*, 490 F. Supp. 968.

[317] *Harris*, 490 F. Supp. at 971.

[318] *Id.*

ployee, the court did not bar the plaintiff's action but questioned whether she would be able in the exercise of due diligence to determine whether the defendant was employed by the government.[319]

The diligence discovery rule requires that a plaintiff make a reasonable effort to determine whether an injury exists and the cause of that injury; in these situations, identification of the tortfeasor is part and parcel to learning the injury's factual cause. One could thus argue that a diligence discovery regime ought to apply to ascertaining the alleged tortfeasor's government-actor status. Courts still may avoid harsh results under such a standard in cases in which plaintiffs are unaware that defendants are government employees and fail to file claims in compliance with the FTCA's administrative time constraint; if the court determines that the plaintiff (1) exercised reasonable diligence to determine the identity of the alleged tortfeasor, and (2) in the exercise of such diligence could not have been reasonably expected to discover that the defendant was a government employee, the court should toll the statute based on equity.[320] Interpreted in this way, the diligence discovery rule itself eliminates the need to make special provision for cases in which the government-actor status of a defendant is unknown to the plaintiff. However, it might be prudent to add the phrase "and the governmental status of the tortfeasor" to the definition of knowledge of the fact of injury required to trigger the running

---

[319] Gould v. United States Dep't of Health & Human Serv., 884 F.2d 785 (4th Cir. 1989). The court determined that the plaintiff had no indication that the defendant physicians, who worked in a private health facility, were employees of the U.S. Public Health Service; thus she had no reason to suspect that her claim was governed by the FTCA. *Id.* at 788. The court reasoned that

> although [plaintiff] Gould was probably aware soon after her husband's death that his death was caused by medical malpractice, she had no way of knowing that the principal causative actor contributing to his death was a government employee. She was, therefore, not "in possession of the critical fact[] . . . [of] who has inflicted the injury," . . . and, before the government informed Gould that [defendant] O'Rourke was a federal employee, did not have any "knowledge to put [her] on inquiry" . . . or any "notice to prompt [her] to explore the legal ramifications of the government's involvement."

Id. at 788 (quoting *Kubrick*, 444 U.S. at 122; *Wilkinson*, 677 F.2d at 1002; and *Henderson*, 785 F.2d at 126). *But see* Flickinger v. United States, 523 F. Supp. 1372 (W.D. Pa. 1981) (dismissing malpractice suit for failure to file an administrative claim within two years of injury despite the fact that the plaintiff did not know that the nurse who treated her was a federal employee until after the two-year period had expired); Lien v. Beehner, 453 F. Supp. 604, 606 (N.D.N.Y. 1978) (stating that "strong equitable considerations notwithstanding, the two-year limitation period of 28 U.S.C. § 2401(b) cannot be tolled or waived").

[320] *See Gould*, 884 F.2d 785; Harris v. Burris Chem., Inc., 490 F. Supp. 968 (N.D. Ga. 1980).

of the statute. As is evident from the discussion above, these difficulties arise with sufficient frequency to suggest that the legislature should err on the side of spelling out the protections (and corresponding duties to investigate) within the diligence discovery approach.

D. *Sunset Provisions*

Congress should also consider whether to include a sunset provision to limit the time in which a tort claim may be brought regardless of when the cause of action was discovered. The Kansas statute includes such a provision, which states that

> if the fact of injury is not reasonably ascertainable until some time after the initial act, then the period of limitations shall not commence until the fact of injury becomes reasonably ascertainable to the injured party, *but in no event shall an action be commenced more than 10 years beyond the time of the act giving rise to the cause of action.*[321]

Many states that have adopted a diligence discovery rule have included a sunset provision.[322] Other states use such provisions in areas other than tort, labelling them statutes of repose because they cut off exposure to litigation risk.[323]

The inclusion of a sunset provision poses inevitable line-drawing dilemmas but affords an opportunity to set a desirable outer limit on the continuation of exposure to liability. Notwithstanding the goal of assuring a just "vindication of the plaintiff's rights"[324] by adoption of a sensible approach to application of the limitation period, concerns about fading evidence and fairness to defendants cannot be ignored. Insurance considerations also may urge inclusion of such a provision.[325]

---

[321] KAN. STAT. ANN. § 60-513 (Supp. 1990) (emphasis added).

[322] *See, e.g.*, CONN. GEN. STAT. ANN. § 52-584 (West Supp. 1990) (three years from act or omission complained of); FLA. STAT. ANN. § 95.11(4)(b) (West 1982) (four years from date of incident out of which action arose; seven years in cases of concealment of injury from plaintiff); HAW. REV. STAT. § 657-7.3 (1988) (six years after act or omission causing injury or death); IOWA CODE ANN. § 614.1(9) (West Supp. 1990) (six years after act, omission, occurrence).

[323] *See, e.g.*, VA. CODE ANN. § 8.01-250 (1984); Eagles Court Condominium Unit Owners Ass'n v. Heatilator, Inc., 239 Va. 325, 389 S.E.2d 304 (1990).

[324] Burnett v. New York Cent. R.R., 380 U.S. 424, 428 (1965).

[325] Yamaguchi v. Queen's Medical Center, 65 Haw. 84, 648 P.2d 689 (1982), offers a thoughtful discussion of the connection between insurance considerations and sunset provision policies. The Hawaii court stated that when the state's diligence discovery rule for medical malpractice actions was adopted, "the legislature was not blind to the rising cost of malpractice insurance and problems of proof attendant with stale claims,

We propose the following amendment to section 2401(b) to incorporate a sunset provision (changes noted in bold print):

(b) Except as provided in subsections (c) and (d) of this section, a tort claim against the United States shall be forever barred unless it is presented in writing to the appropriate federal agency within two years after such claim accrues or unless action is begun within six months after the date of mailing, by certified or registered mail, of notice of final denial of the claim by the agency to which it was presented. A tort claim against the United States shall be deemed to have accrued [insert language of Alternatives A, B, or C], **but in no event shall any action be commenced more than ____ years after the date of the alleged act or omission causing the injury** [or death]. **This ____-year time period shall be tolled for any period during which it can be shown that fraud, concealment, or intentional misrepresentation of fact prevented the discovery of the injury** [and/or its cause in fact] **[by a plaintiff** [claimant] **in the exercise of reasonable** [due] **diligence].**[326]

This draft provision draws upon Hawaii's sunset provision and Florida's statute tolling the limitation period for medical malpractice actions in which fraud, concealment, or intentional misrepresentation prevented plaintiff's discovery of an injury or its cause.[327] It balances the need to have an absolute limitation period for claims that are governed by a diligence discovery rule against the need to account for cases in which the defendant's conduct itself prevents a plaintiff from discovering the injury and its cause in fact.

---

both of which were subject to exacerbation under an open-ended 'discovery rule.'" *Id.* at 88, 648 P.2d at 692. The court cited the report of the legislature's judiciary committee on the proposed six-year limitation, which noted that malpractice insurance premiums increased under a discovery rule because "insurers were required to hold in reserve funds for an indeterminable period of time." *Id.* at 88 n.9, 648 P.2d at 692. The report recommended a six-year limitation on claims because "[t]estimony presented indicated that this [limitation period] would have a tendency of lowering the cost to doctors of maintaining malpractice insurance in that insurers could hold their reserves for a fixed period of time . . . ." *Id.* (quoting HOUSE STAND. COMM. REP. NO. 455, 7TH HAW. LEG., 1ST SESS., *reprinted in* HOUSE JOURNAL 947 (1973)).

326 Current sunset provisions in states that have a diligence discovery rule range from three years to 10 years. *See, e.g.*, CONN. GEN. STAT. ANN. § 52-584 (West Supp. 1990); KAN. STAT. ANN. § 60-513 (Supp. 1990); N.C. GEN. STAT. § 1-52(16) (1983); OR. REV. STAT. § 12.115 (1989).

327 *See* HAW. REV. STAT. § 657-7.3 (1988); FLA. STAT. ANN. § 95.11(4)(b) (West 1982).

V. Conclusion

Several models exist for drafting an amendment to the limitation provision of the FTCA. These include the limitation rule for actions brought *by* the United States, as well as limitation sections in other statutes, in draft bills previously introduced, and in the Restatements.

Given the support identified in the case law for the diligence discovery approach, and the feasibility of amending section 2401(b) in a fairly simple fashion to achieve the goals of greater uniformity and fairness in court decisions, as well as to reduce the costs associated with private bills and fulfill the remedial purpose of the FTCA, we recommend an amendment along the lines proposed in this Article.

ARTICLE

SCHOLARSHIPS AND THE FEDERAL INCOME TAX BASE

CHARLOTTE CRANE*

Since the passage of the Tax Reform Act of 1986, only scholarship payments to degree-seeking students for tuition, fees, and required books and supplies remain tax exempt.

In this Article, Professor Crane examines the rationale for an income tax exemption for scholarships, focusing on the fact that including scholarships in the income tax base is inappropriate because scholarship recipients may not receive actual value at the time they receive the scholarship and scholarships may simply equalize endowments between individuals. She argues that even if scholarships should be included in the tax base, they should nevertheless be given preferential treatment in the Tax Code. After an in-depth analysis of the nature of the income tax base and tax preferences, Professor Crane concludes that an exemption for scholarships is justified.

In 1986 Congress dramatically limited the kinds of student financial aid that are exempt from taxation.[1] Under prior law, any payment made to further the education of a degree-seeking student was tax exempt, so long as the payment was not made in exchange for services rendered. As a result of the 1986 changes, however, only those scholarship payments for the tuition, fees, and required books and supplies of degree candidates are exempt.[2]

The 1986 reforms were made under the general theme of broadening the tax base,[3] fueled in part by a genuine interest in

---

* Professor of Law, Northwestern University School of Law. A.B., Radcliffe College, 1973; J.D., University of Michigan Law School, 1976. This Article is an outgrowth of a project initiated by the Teaching Taxation Committee of the Tax Section of the American Bar Association and benefited immeasurably from the comments of the members of that group. Special thanks are also due to Michael Durst, Myron Grauer, Karen Glen, Lisa Kuo, and the Cranston and Catherine M. Spray Fund.

[1] Tax Reform Act of 1986, Pub. L. No. 99-514, § 123(a), 100 Stat. 2085 (codified as amended at 26 U.S.C.).

[2] I.R.C. § 117(a) (1988). Unless otherwise noted, all future references to the Internal Revenue Code can be found in the 1988 United States Code.

[3] Despite the tendency to associate base-broadening with simplification, the 1986 changes had clearly identifiable complicating effects. In every case, the most persistent problem under prior law, determining whether payments made are associated with services, remains. Moreover, this determination may not be any easier under the post-1986 laws. *See infra* note 101. Additionally, many new issues will arise concerning how a recipient's required expenditures should be calculated when determining what amounts qualify for exclusion, and to which academic year the payments are related. *See infra* note 79.

Furthermore, the possibility of a taxable payment qualifying as a scholarship opens up whole new areas of confusion. A payor has no payroll tax or reporting obligations reform[4] and in part by a need to raise revenue.[5] The changes in the treatment of scholarships were justified in the legislative history largely as efforts to minimize the difference between the tax burden placed on those who must pay for education from their individual earnings and the tax burden on those who receive scholarship grants.[6] The early proposals underlying the changes suggested that even the retained exemption for amounts used to pay tuition could be justified only in terms of the hardship on the recipient if such amounts were taxed. In this view, political expediency may be the only reason that tuition scholarships are still exempt.[7]

---

with respect to taxable scholarships; however, it does have reporting requirements for payments to non-employees, and it has both reporting and payroll tax obligations with respect to payments to employees. *See infra* note 103.

To complicate matters further, there may be a distinction between payments to current and future employees with respect to these obligations. *See* Letter from James W. Quiggle to Commissioner of the Internal Revenue Service (August 12, 1988), *summarized in* TAX NOTES, Aug. 15, 1988, at 681.

Finally, the Code formerly had assumed that any amount characterized as a scholarship would be exempt from taxation for the recipient. Several provisions cannot easily be applied now that this is no longer true. Examples include: the provisions of I.R.C. §§ 4941 & 4945 that refer to scholarships subject to § 117(a) (addressed in I.R.S. Notice 87-31, 1987-1 C.B. 475); the treatment in § 63(c)(5) of a child's earned income including taxable scholarships; and the exclusion of scholarships in § 152(d) when determining support for dependency.

4 The inclusion of scholarships in the tax base has been listed as an achievement of the 1986 Act by at least one major supporter of base-broadening. *See* Pechman, *Tax Reform: Theory and Practice*, 1 J. ECON. PERSP. 11, 18–19 (1987).

5 It is unclear whether the scholarship provisions themselves were the subject of much revenue-driven pressure. Despite the relatively high numbers used when considering scholarships as a tax expenditure, *see infra* note 17, the 1986 Act changes were estimated to provide $8 million in revenue for 1987, $64 million for 1988, $130 million for 1989, $160 million for 1990 and $164 million for 1991. STAFF OF JOINT COMM. ON TAXATION, GENERAL EXPLANATION OF THE TAX REFORM ACT OF 1986, at 1359 (Comm. Print 1987) [hereinafter EXPLANATION OF ACT]. Much of the disparity results from the fact that only part of the exemption was removed; the rest is explained by the substantial amount of grandfathering. All scholarships "granted" before the 1986 enactment, regardless of the year to which they related, were grandfathered. I.R.C. § 151(d).

6 The 1984 Treasury proposals stated that an exclusion for scholarships was unjustified because it was "unfair to the ordinary taxpayer who must pay for education with earnings that are subject to tax." U.S. TREASURY DEP'T, TAX REFORM FOR FAIRNESS, SIMPLICITY AND ECONOMIC GROWTH—GENERAL EXPLANATION OF TREASURY DEPARTMENT PROPOSALS 58 (1984) [hereinafter EXPLANATION OF PROPOSALS].

The House Report used this argument in a slightly different way to explain the change, distinguishing tuition from room and board scholarships:

> [Prior law] provides a tax benefit [in providing an exclusion for room and board] not directly related to educational activities; by contrast, students who are not scholarship recipients must pay for such services out of after-tax dollars.

H.R. REP. No. 426, 99th Cong., 1st Sess. 100 (1986). The General Explanation used essentially the same language, again emphasizing the disparity in tax treatment between those who use their own funds and those who receive scholarships. EXPLANATION OF ACT, *supra* note 5, at 40.

7 The EXPLANATION OF PROPOSALS states:

Criticism of the scholarship exemption is not new.[8] Some commentators have objected to the scholarship exemption on grounds similar to those stated in the 1986 legislative history; they argue that scholarships represent wealth accretions just as much as any other payment.[9] Other commentators have repeatedly objected to the way the provisions operate in a progressive system to provide more benefit to higher-bracket recipients than to lower-bracket recipients.[10]

Although these criticisms of a scholarship exemption have some merit, they fail to present a complete analysis in three respects. First, it is impossible to create a perfectly "comprehensive" tax given that wealth and well-being cannot always be reduced to monetary terms. Education is among the human endeavors least amenable to this reduction. An exemption for educational assistance can be seen as mitigating some of the inherent limitations in any definition of a comprehensive tax base. Second, even if an exemption for scholarships can be justified only in terms of a preference and not in definitional terms, it seems likely that it is a far more desirable preference than most. Third, the kind of education typically subsidized

---

> In theory, it might be appropriate to include the full amount of any scholarship in income. In practice, this would create real hardships for many scholarship recipients. Scholarship awards are often made on the basis of need, and if students were taxed on such amounts, they would often not have the resources to pay the tax.

EXPLANATION OF PROPOSALS, *supra* note 6, at 58–59. This language was repeated in the President's Proposals to the Congress for Fairness, Growth, and Simplicity, Public Papers of the Presidents Weekly Compilation of Presidential Documents, 1985 U.S. CODE CONG. & ADMIN. NEWS 57.

8 Interestingly, various comprehensive base-broadening proposals have reached different conclusions on the appropriate treatment of scholarships. The Carter Commission in Canada, which otherwise espoused a dramatically broad view of the individual income tax base, advocated a credit for tuition and the living costs of post-secondary students, apparently on the ground that it was an appropriate incentive measure. Implicit in the proposal is an exemption for scholarship receipts not involved in the calculation of the credit. 3 CARTER COMMISSION, THE REPORT OF THE ROYAL COMMISSION ON TAXATION 236–37 (1966).

On the other hand, proposals have been made to include all cash transfer payments in income. *See, e.g.*, D. BRADFORD, BLUEPRINTS FOR TAX REFORM 57–58 (1984) [hereinafter BRADFORD]. Although scholarships are not specifically mentioned in the discussion of transfer payments, their inclusion was clearly intended. For instance, in the sample tax form provided in the text, scholarships are listed as examples of "public assistance benefits." *Id.* at 127.

9 *See, e.g.*, U.S. COMM. TO REVISE THE TAX STRUCTURE, REFORMING THE FEDERAL TAX STRUCTURE 21 (1973).

10 *See, e.g.*, 1 B. BITTKER, FEDERAL INCOME TAXATION OF INCOME, ESTATES AND GIFTS 11–19 (1981). Others have focused their criticism on the fellowship exemption for non-degree candidates. *See, e.g.*, Wolfman, *Federal Tax Policy and the Support of Science*, 114 U. PA. L. REV. 171, 186 (1965). As discussed below, *see infra* note 108, these criticisms rest on very different grounds.

under the old rules is only one kind of human capital development; other kinds of development are still treated favorably under the tax law. Eliminating the exemption for scholarship receipts might in fact eliminate the least objectionable of these existing tax benefits.[11]

This Article examines the support for scholarship exemptions in greater depth than most of the arguments that would have been advanced in favor of the initial exemption for scholarships. The exemption originated in the relatively simple government position that scholarships were excludible gifts, and the codification of the exemption in 1954 seems to have relied primarily on this gift aspect of scholarships.[12] The questions raised before this rule was codified focused on the problems created when a student's services were related to the grant.[13]

---

[11] In outlining these arguments, this Article focuses only on educational assistance that is student-directed. It ignores the assistance received by the institution itself, including the exemption from income tax enjoyed by educational institutions and the treatment of charitable donations which they recieve.

[12] The American Law Institute proposal excluded scholarships in its general treatment of "awards, prizes, scholarships and similar payments," when they were "made to further the pursuit by a student of educational, vocational, or similar activities, and not as compensation for services." 32 FEDERAL INCOME TAX STATUTE, § x107(m) (Draft 1954) [hereinafter ALI STATUTE]. This draft focused upon the problems encountered under the general exclusion for gifts contained in prior law, for the explanation merely states that "all the usual student fellowships" would be excluded. ALI STATUTE, *supra*, Comment on § x107(m), at 207.

H.R. 8300, which became the 1954 Code, separated the provisions for scholarship and fellowship grants from those for other awards. H.R. 8300, 83d Cong., 2d Sess. (1954). But the legislative history from this point on is concerned only with resolving questions about service-related payments, and not with justifying the exemption. H.R. REP. No. 1337, 83rd Cong., 2d Sess., pt. 1, at 17, A37 (1954); S. REP. No., 83rd Cong., 2d Sess. 189 (1954).

[13] *See* I.T. 4056, 1951-2 C.B. 8, *rev'd*, Rev. Rul. 69-43, 1969-1 C.B. 310.

Shortly after the enactment of the 1954 Code, however, the Service asserted that all "amount[s] paid or allowed to, or for the benefit of, a student . . . to aid such individual in pursuing his studies" were scholarships, subject to Treas. Reg. § 1.117-3(a), and that therefore the exemption provided in I.R.C. § 102 for gifts was inapplicable. Treas. Reg. § 1.117-1(a). It excluded from the definition of scholarship "any amount provided by an individual to aid a relative, friend, or other individual in pursuing his studies where the grantor is motivated by family or philanthropic considerations." Treas. Reg. § 1.117-3(a). Given the plausibility of the argument that the purpose of I.R.C. § 117 was to ensure the exemption of educational assistance that might arguably have been connected to services and thus not considered a gift, the Service's position could be seen as overreaching. Furthermore, the distinction in the regulation between transfers made by natural persons and transfers made by other entities is peculiar, except as an overgeneralization reflecting the likelihood of the necessary donative intent.

There seems to have been little pressure on these positions perhaps because, until 1986, the difference between exemption as a gift and exemption as a scholarship was unlikely to make a difference. *Cf.* Cass v. Commissioner, 86 T.C. 1275 (1986) (interpreting the award and scholarship provisions as being mutually exclusive and thus concluding that an award based upon prior achievements but designed to further the

A good and sufficient rationale for eliminating the exemption in 1986 would have been that it no longer served the purpose Congress intended it to serve in 1954. It could not, since the world of scholarships had changed so dramatically.[14] In 1954, most scholarships were given by local business leagues, charitable organizations, and educational institutions themselves. Recognition of their affinity with gifts generally, mixed perhaps with a doubt that transfer payments should be included in income, adequately explained the initial exemption.

In the ensuing thirty-five years, however, these forms of scholarships have become almost trivial compared with the massive amounts of scholarship aid the federal government provides. Although some doubts about the propriety of including any transfer payments in the definition of income persist to this day,[15] the analysis of the scholarship exemption must focus upon the proper role of the tax system in implementing other social programs. Ironically, the change in the nature of scholarship distribution, from completely haphazard autonomous programs to a more coordinated set of programs involving enormous federal participation and increasing federal control, seems to enhance rather than reduce the justification for an exemption.

Uncertainty about the proper treatment of scholarships is evident from their treatment as tax expenditures.[16] Until 1983,

recipient's future research could qualify solely as a fellowship for which only partial exclusion would be available); Isenbergh v. Commissioner, 31 T.C. 1046 (1959).

The regulations proposed after the 1986 legislation repeat this position. *See* Prop. Treas. Reg. § 1.117-6(c)(3), 53 Fed. Reg. 21,688 (1988). After 1986, there is likely to be far more pressure on the lines distinguishing "gifts" that remain exempt and gratuitous transfers made with the particular purpose of aiding the recipient's education that are not exempt. For instance, suppose the local garden club gives a $500 "scholarship" to a deserving high school graduate planning to attend college, but without any restriction on how it can be spent; can this qualify as a gift? Is it includible as an award? And what if the recipient's grandmother is a member of the club? Is that more or less likely to make the transfer exempt? Does it matter whether contributions to the club were themselves deductible?

[14] There was no broad-based federal scholarship aid until passage of the National Defense Education Act of 1958. Moreover, federal student aid, not counting subsidized loans, went from 27% of total tuition costs in 1964 to 76%, if guaranteed loans are counted, in 1986. B. BOSWORTH, A. CARRON & E. RHYNE, THE ECONOMICS OF FEDERAL CREDIT PROGRAMS 128 (1987) [hereinafter BOSWORTH]. For a general history, see C. FINN, SCHOLARS, DOLLARS AND BUREAUCRATS 59–67 (1978) [hereinafter FINN].

[15] *See, e.g.*, Lane, *A Theory of the Tax Base: The Exchange Model*, 3 AM. J. TAX POL'Y 1 (1984); Warren, *Would a Consumption Tax Be Fairer Than an Income Tax?*, 89 YALE L.J. 1081, 1085–86 (1980).

[16] Allowance of a tax reduction implicitly subsidizes a deductible or excludible expenditure or receipt. Implicit subsidies of this type are referred to as tax expenditures and are required to be estimated as part of the government's budgeting process. W. ANDREWS, BASIC INCOME TAXATION 18 (3d ed. 1985).

the exclusion for scholarships was listed as a tax expenditure,[17] but, beginning in 1983, the status of the scholarship exemption was changed with the substitution of the "reference tax" concept for the "normal tax" concept.[18] The reason, as later stated, was that:

> from a strictly economic point of view, scholarships . . . are either gifts not conditioned on the performance of services, or they are "rebates" of educational costs by the institutions in which students are enrolled. Thus . . . the exclusion is not a tax expenditure; the reference law does not include either gifts or price reductions in a taxpayer's gross income.[19]

Despite this change in rhetoric, the estimates made under the old rationale have been retained in the analysis for eight years. The estimate for 1988 included in the 1990 budget was $570 million, and was expected to rise to $655 million in 1990.[20]

I. Educational Assistance and a Broad-Based Income Tax

Even purists agree that the federal income tax base is not likely to encompass all measurable accretions of wealth. The income tax's reliance on the establishment of market prices, and the related limitations associated with realization, preclude the possibility of such a comprehensive tax base. Because these

---

[17] This listing appears in Special Analysis G, appended to the Budget of the United States Government. The value of this exclusion over the last decade has ranged from $245 million, estimated for 1977 in the 1979 budget, Office of Management and Budget, Budget of the United States, Fiscal Year 1979, at 159 (Table G-1), to $645 million estimated for 1986 in the 1986 budget, Office of Management and Budget, Budget of the United States, Fiscal Year 1986, at G-45, and $995 million for 1987 in the 1987 budget, Office of Management and Budget, Budget of the United States, Fiscal Year 1987, at G-44.

[18] The number of items counted as tax expenditures was reduced by this change in the baseline for determining tax expenditures. *See generally* Office of Management and Budget, Budget of the United States, Fiscal Year 1983, at G-8. As explained in the Budget for fiscal year 1986, the "normal tax" concept was close to the concept of a comprehensive tax base. Office of Management and Budget, Budget of the United States, Fiscal Year 1986, at G-2. Using the "reference tax" as a baseline resulted in the inclusion as a tax expenditure of only those items for which there was some inconsistent general rule and for which an equivalent grant program could be designed. *Id.* at G-5.

[19] Office of Management and Budget, Budget of the United States, Fiscal Year 1985 at G-28.

[20] Office of Management and Budget, Budget of the United States, Fiscal Year 1990, at G-52. The cost of direct expenditure programs designed to produce the same results was $720 million. *Id.* at G-42.

inherent limitations do not always reflect consciously espoused social values, care must be taken that they do not give rise to arbitrary and distortive distinctions among sources of well-being.

Three distinct arguments suggest that inclusion of tuition scholarships in an income tax base subject to such constraints would be inappropriate. First, given the realities of the higher education market, a scholarship recipient has not necessarily received any particular amount at the time he receives the benefit of the scholarship,[21] especially when he is compared to others receiving education subsidized in other ways. Second, recipients of needs-based scholarships should not be taxed because these scholarships in effect replace values already enjoyed without taxation by nonrecipients. Third, even though the scholarship recipient is incidentally receiving some personal consumption-like benefit at the time he receives the scholarship, the extent of this benefit is sufficiently uncertain to make taxation inappropriate.

These arguments are of the sort frequently included in attempts to define the tax base and are therefore presented as such here.[22] They can be distinguished from those arguments justifying exceptions that can only be made after certain basic definitions have been established. All of these arguments, however, could be presented to support a scholarship exemption, whether as part of the basic definition of the income tax base or not.

A. *Educational Assistance: In-kind, In-credit, and In-cash*

One limitation on the income tax base is its reliance on markets to make determinations of value. In general, if there is no independent market for an in-kind benefit, its receipt probably will not be viewed as income. For instance, the types of in-kind benefits traditionally provided primarily by governments have

---

[21] *See infra* text accompanying notes 52–53.

[22] *See, e.g.*, Andrews, *Personal Deductions in an Ideal Income Tax*, 86 HARV. L. REV. 309 (1972). There is a sense in which the arguments in this Article differ from those presented by Professor Andrews. Professor Andrews attempted to set out a definition of income, focusing on the recipient's personal consumption and against which all transactions could be measured, without regard for any systemic limitations on the ability to implement the definition. Much of the following text, however, is based on the notion that there can be no such definition without taking into account the limitations on attempts to implement any definition.

never been included in income.[23] Municipal services, such as police and fire protection, library privileges, public recreation, and garbage removal, have never been included in the income tax base. This is true even when the service is not provided to all citizens, and even when these same services are incidentally provided to others outside the taxing jurisdiction on a fee-basis. Similarly, there is no attempt to isolate an item of income when value is transferred in other contexts at prices that are not subject to market forces for other reasons and that reflect heavy subsidies. For example, when one pays only a nominal amount for admission to a museum, there is no attempt to include the "bargain element" of the admission price in income.

For the most part, ignoring these types of values does no serious harm. It creates neither inequity nor economic distortion. Moreover, excluding all such benefits avoids vexing questions about valuation, particularly about whether those who do not partake of the government services available to them should be held to have income merely because the service was available. More important, for many services provided by government, everyone has an equal opportunity to benefit from them; therefore, excluding these services from the income of those who do partake creates no serious inequities for those who do not. Even if everyone cannot partake, with the result that an exemption cretes some inequity, the exemption will still only cause a minimal market distortion so long as the services are provided by similarly situated institutions using the same general approach to pricing and funding.

---

[23] *See generally* Aaron, *What is a Comprehensive Tax Base Anyway?*, 22 NAT'L TAX J. 543 (1969); Bittker, *A "Comprehensive Tax Base" as a Goal of Income Tax Reform*, 80 HARV. L. REV. 925, 936 (1967). Ironically, the latter author chose education subsidies to demonstrate the fallacy of an ideal comprehensive tax base, but the author failed to note that this particular subsidy, because of the decentralized and varying way in which it is administered, is in fact even more problematic than most other government subsidies.

The government reckoning of tax expenditures reflects a similar approach:

> The exclusion from gross income of direct cash payments to individuals by the Government, such as social security payments, does result in tax expenditures. Other Government programs extend benefits in kind to individuals. Examples are medicare and public education. Since these benefits are received in kind they cannot be used, like cash, for purposes fully consistent with the recipient's preferences. Moreover, their cash value is often difficult to identify with certainty. Thus the exclusion of in-kind benefits from income subject to tax is not considered to result in tax expenditures. The dividing line between nontaxable Government benefits that do result in tax expenditures and those that do not is essentially arbitrary.

OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES, FISCAL YEAR 1978, at 151.

This analysis may justify and explain the tax-exempt status of fire and police protection, public transportation, and perhaps even primary education. The institutions providing higher education, however, are not homogeneous, either in their funding techniques or in their approaches to pricing. Further analysis is necessary to explain the exemption in this area.

Virtually all higher education is subsidized in the sense that tuition paid by individual students out of their own and family resources does not cover the entire cost of running the institution and providing the educational environment from which the students benefit.[24] However, the ways in which this shortfall is made up are not at all uniform. In public institutions, the subsidy is in large part provided out of tax dollars, as tuition is inadequate to cover even the direct costs of an individual student's education. In private institutions, much of the subsidy comes from private contributions, and even though it may be harder to prove the subsidy for any particular student, it still exists.[25] In both cases, determining the full extent of the subsidy is difficult.

No one would suggest that this subsidy should be taxable income to the student. However, the legislative history of the 1986 Act suggests that, to the extent tuition is actually charged and not collected, the student should in theory have taxable income. The resulting distinction between the invisible subsidy inherent in low stated tuition rates and the visible subsidy in-

---

[24] Bok, *What's Wrong with Our Universities?*, HARV. MAG., May–June 1990, at 44, 51.

[25] The DIGEST OF EDUCATION STATISTICS for 1987 reports that in 1984–85 the expenditure per pupil at public four-year colleges was $11,330, while the corresponding amount at private four-year colleges was $14,963 in 1984–85. NATIONAL CENTER FOR EDUC. STATISTICS, DIGEST OF EDUCATION STATISTICS 235 (1987) [hereinafter EDUCATION STATISTICS]. Total charges to students at four-year colleges affiliated with universities were estimated to be $4,370 for public institutions and $11,870 for private institutions in 1986–87. *Id.* at 223. These figures suggest that public school students face a stated tuition of less than 40% of the actual cost of their education, while private school students face a stated tuition of almost 80% of the cost. However, these figures are only rough approximations, because, among other reasons, the total expenditure figures include both the expenditures of university-affiliated and independent four-year colleges, while the tuition figures are only for university-affiliated institutions.

The potential disparity between the treatment of cash subsidies and in-kind subsidies becomes even more apparent in light of the fact that, in 1985, 5,210,000 students were enrolled in four-year public institutions and only 2,463,000 were enrolled in four-year private, non-profit institutions. NATIONAL CENTER FOR EDUC. STATISTICS, THE CONDITION OF EDUCATION: A STATISTICAL REPORT 120 (1987). This same report indicated that, in 1985, public institutions derived 18% of their revenues from tuition, while private institutions derived 55% of their revenues from tuition. *Id.* at 116.

herent in high tuition rates accompanied by generous scholarship funding is difficult to justify.[26]

Including the entire tuition charged but not collected in income would produce very little scholarship income for those attending most state institutions. When artificially low tuitions are charged to everyone regardless of ability to contribute more, no student is deemed to have received a scholarship. Such a rule, however, would produce inordinate amounts of income for students attending those private institutions that have raised tuitions to reflect the real ability of some students to pay. When nominally high tuitions are charged but are actually paid only by a small fraction of the students attending, the other students would be deemed to have income.

These tuition practices of private institutions amount to price discrimination, through which individuals pay what they are willing and able to pay for the services provided. These institutions attempt to force financially able students to bear as high a portion of their costs as possible, and provide generous scholarships to a substantial number of less able students. Thus no single "market value" of the education being received exists.

Such practices are inconsistent with many of the operating assumptions that make an income tax attractive and possible. The income tax requires that an item have a fair market value, a single price that adequately reflects aggregate, if not individual, preferences. Subjective values and consumer surplus are ignored.[27]

This approach to value is simply inadequate to deal with the price discrimination prevalent in the higher education market. The idea that people are paying different amounts for exactly the same product under exactly the same circumstances is inimical to the income tax notion of fair market value. If one insisted on including educational assistance in income, it does

---

[26] This dilemma—the effect of taxing cash transfer payments, which are often distributed on a needs-basis, but not taxing in-kind benefits, which are less frequently needs-based—has been noted elsewhere. *See, e.g.*, Sunley, *Employee Benefits and Transfer Payments*, in COMPREHENSIVE INCOME TAXATION 103 (J. Pechman ed. 1977). The unique aspect of this dilemma in the case of scholarships, however, is the presence of both forms of transfer in direct competition and outside the government's control.

[27] Different substantive rules have been created for those situations in which this assumption seems harsh, such as in-kind employee benefits and gifts. The rule used most often in such cases is to exclude the income, with the result being an "all or nothing" approach to valuation in the income tax generally. *See generally* Crane, *Matching and the Income Tax Base, The Special Case of Tax Exempt Income*, 5 AM. J. TAX POL'Y 191, 232 (1986) [hereinafter Crane].

not make sense to try to measure it by the difference between the nominally charged tuition and the amount actually paid by the taxpaying student. Moreover, if one insisted that the receipt of the scholarship should be taxable, either the difference between the institution's cost per pupil and actual amount paid, or the present value of the additional lifetime earnings anticipated as a result of the education would be far more sensible measures of the economic benefit to the student. These measures of income, however, do not reflect traditional notions of income for income tax purposes.

A rule that created income when a high tuition was actually charged would affect decisions about how to subsidize education. Such a rule would clearly favor the tax-supported public provision of education over the tuition-supported and gift-supported private provision of education. Any subsidy, whether public or private, that allowed the institution to lower tuitions for all students would be favored over subsidies that were directed at individual students.[28] Furthermore, the rule would tend to promote the establishment of subsidized state institutions rather than the distribution of the cash subsidy from the state to individual students for use at an institution of their choice.[29] This rule would reduce even further what little market pressure exists in higher education, and would make the assistance less likely to further other governmental goals, including endowment replacement.[30]

Such a rule would also frustrate price discrimination based upon ability to pay. Each dollar of higher stated tuition collected from the financially able would result in the imputation of a

---

[28] Some preference for institution-directed aid over individual-directed aid may already exist. For example, a charitable deduction is generally available only when payment is made to an institution, not when it is made to an individual. *See, e.g.*, Rev. Rul. 68-484, 1968-2 C.B. 105. Additionally, under I.R.C. § 4945, gifts of private foundations to individuals are more likely to be subject to penalty tax. These rules, however, serve the broader purpose of preventing abuses of the charitable deduction and exempt status and do not reflect any overall preference for institution-directed aid.

[29] Some states have consciously attempted to offer the student a choice between an in-kind subsidy at a state school and a cash subsidy to be used at a private school. *See* Hearn & Anderson, *Integrating Postsecondary Education Policies: The Minnesota Model*, in STUDYING THE IMPACT OF STUDENT AID ON INSTITUTIONS 55 (R. Fenske ed. 1989).

[30] *See infra* text accompanying note 33. Some writers assume that any institution-directed aid will be more regressive than student-directed aid. *See, e.g.*, BOSWORTH, *supra* note 14, at 131. There is evidence that subsidies to state institutions are already one of the most regressive forms of student aid. *See* McPherson, Schapiro & Winston, *The Impact of Federal Student Aid on Institutions: Toward an Empirical Understanding*, in STUDYING THE IMPACT OF STUDENT AID ON INSTITUTIONS, *supra* note 29, at 31.

dollar of taxable income to the financially less able. The institution might well abandon its efforts to charge as much as its most affluent students could pay when, for every dollar redistributed in this way, the federal government claimed fifteen cents. A rule that did tax scholarships when an institution charged a high tuition and provided tuition credits, but did not tax in-kind benefits provided without such bookkeeping entries, would put substantial pressure on the institution to provide the benefits to all individuals as in-kind benefits by charging well below what the most able student would be willing to pay.[31]

In sum, two related aspects of educational assistance make its inclusion in the income tax base problematic. First, much assistance is provided in-kind as a government subsidy; second, much assistance is distributed through a market in which price discrimination is prevalent. The problems of taxing in-kind assistance, such as government transfers, have simply been ignored in most efforts to define the tax base. Furthermore, the problems relating to price discrimination have not been fully considered, since markets in which effective price discrimination has prevailed are rare.

It is unlikely that either of these related problems in defining the income tax base can be overcome. Therefore, it makes sense to try to develop a definition of the tax base that provides consistent treatment of equivalent items.[32] Thus, an exclusion for tuition scholarships can be justified as an effort to equalize the treatment of those who receive in-kind benefits, either from public institutions charging low tuitions or from private institutions that subsidize their programs, and those who receive cash-equivalent benefits in the form of credits against tuition charges.

---

[31] There is no universal support for needs-based scholarship distribution and the resulting pattern of price discrimination. Some criticize the openly redistributive effects of such a system as well as the lack of market discipline and consequent inefficiencies that can result from the fact that significant costs are not borne by the actual consumer. *See, e.g.*, Carne, *The Campus Cost Explosion*, 40 POL'Y REV. 68 (1987). Another writer has speculated that the mechanisms needed to implement such a system for price discrimination result in opportunities for other, more nefarious pricing behavior. *See* Ayres, *Colleges in Collusion*, NEW REPUBLIC, Oct. 16, 1989, at 19.

[32] This argument has been adopted here despite the potential criticism that it is facile or somehow unprincipled. *See* Griffith, *Theories of Personal Deductions in the Income Tax*, 40 HASTINGS L.J. 343, 370 n.153 (1989). Equalizing by expanding an exclusion may be appropriate when there is some doubt about the appropriate treatment of both items and little can be done about the treatment of one. Of course, equalizing by expanding an exclusion will not be appropriate when the excluded item clearly should be included, and when it is just as feasible to include the item as it is to exclude the item. *See infra* text accompanying note 78.

B. *Educational Assistance as Replacement for Endowment*

1. Endowment in the Form of Cultural Environment

Even the most comprehensive income tax can never fully tax the values an individual enjoys from the happenstance of her individual circumstances. These untaxed advantages are subtle and virtually impossible to quantify. Some people have parents that provide a stimulating and challenging home environment, others have grandparents who take time out to enrich their grandchildren's lives, and still others live in an area where educational expectations are generally high. Similar advantages persist even in young adulthood as some individuals may be able to find summer jobs with their mothers' employers, or to learn through the outreach efforts of local civic organizations. No one is subject to tax on this economic endowment as she receives it, or even as she uses it in the course of transforming it into marketable skills. Instead, this potential economic advantage is only taxed as it is brought into the marketplace and, literally, cashed in.

To the extent that scholarships fund education that serves as a replacement for at least some of these advantages, they are merely providing to some individuals what other individuals have been able to enjoy without being taxed. Excluding such amounts from taxation would treat those individuals receiving scholarships equal to those receiving educational advantages from their environment. Like many kinds of general welfare payments distributed by governments, these payments could be excluded in a way that is consistent with views of the appropriate income tax base.[33]

This justification for a tuition exemption is only plausible if most scholarship aid is in fact distributed on a needs-basis and if the measure of need used actually corresponds to a lack of endowment. This justification does not rely upon the idea that the income tax must be progressive or upon the idea that anything that enhances its progressivity must be accepted as a proper part of the definition of income. Instead, it relies upon the premise that, regardless of how progressive other features

---

[33] A surprising number of these exclusions have been treated as tax exempt receipts even without statutory authorization. For a catalog, see Crane, *supra* note 27, at 233–36.

of the tax are, the tax base can properly be defined to exclude values that are merely replacements for values that others can enjoy without being taxed.

Whether the existing scholarship distribution system can be characterized as such an endowment replacement system is uncertain.[34] First, although most potentially taxable scholarship aid is provided at least in principle on a needs-basis,[35] there is evidence that the number of non-needs-based scholarships is increasing.[36]

Second, the current formulas may not adequately test for need. The federal formulas, which focus on funds currently available rather than on total financial assets, suggest the merit of this criticism. For instance, life insurance, pensions, IRA's, and similar assets are all excluded from assets available for education, despite varying degrees of access to such funds.[37] The family that has saved with retirement in mind without taking advantage of such savings vehicles must include their retirement savings in assets available for education. Any needs test relying on such a formula will obviously allow more aid for middle class students than would be available if a more complete financial

---

[34] John Lee reports that such noble principles are only faintly reflected in the total subsidies to higher education. These principles are only slightly more apparent in total student aid, the component of the subsidies most likely to be included in taxable income. *See* J. Lee, The Equity of Higher Education Subsidies (paper for National Center for Postsecondary Governance and Finance, University of Maryland). The average student from a household with income under $7,000 received $1,262 in aid in 1983, while the average student from a household with income over $38,000 received $795. *Id.* at 14. Part of the reason for this result is that, in general, the higher the tuition, the higher the subsidy provided from all sources. If the affluent student is more likely to attend a higher cost institution, her need will be greater, and the absolute amount of aid thus will be higher.

[35] Even if most aid is distributed nominally on a needs-basis, the result may not adequately reflect need. The available data is decidedly inconclusive. In 1981–82, 68% of those attending public institutions, and 83% of those attending private institutions, who came from families with annual incomes under $12,000 received some grant aid; the corresponding figures for students coming from families with incomes over $25,000 were 24% and 37%. EDUCATION STATISTICS, *supra* note 25, at 225. These statistics are based on a survey of high school seniors in 1980. Since the figures are based on the percentage of students receiving grants with no indication of the amount of the grant, token awards of little real financial consequence are reflected the same as full scholarships. Furthermore, these figures provide no insight into whether annual income as reported might reflect need.

[36] *See, e.g.*, Gladieux, *The Issue of Equity in College Finance*, in THE CRISIS IN HIGHER EDUCATION 79–80 (J. Froomkin ed. 1983) [hereinafter THE CRISIS IN HIGHER EDUCATION]; Ehrenberg & Sherman, *Optimal Financial Aid Policies for a Selective University*, 19 J. HUM. RESOURCES 202 (1984) (providing a sophisticated mathematical model that justifies offering highest aid packages to highest ability students, because such aid is likely to have the largest effect on their actual enrollment).

[37] 20 U.S.C. §§ 1070a-6 & 1087vv(g) (1988).

assessment were used. Furthermore, the typical formulas focus only on the past year's income, and not on prior years' income.[38] They thus completely ignore forgone opportunities to save.

Third, the kind of need measurement used may not be an adequate measure of lack of endowment. The recipient of a needs-based scholarship may be no more likely to be in need of a "cultural" subsidy in the form of a scholarship-financed education than a nonrecipient. Children of clergy and academics, for instance, may be more likely than most to be financially needy, but less likely than most to be lacking in most other senses.

This justification for excluding scholarships is therefore inevitably incomplete. Not all of those receiving scholarships are disadvantaged in their environments, and certainly not all of those disadvantaged in their environments receive scholarships. Nevertheless, the more scholarships are distributed on a true needs-basis, the more justifiable their exemption is on these grounds.

2. Endowment in the Form of Financial Support

What if there were no correlation between cultural endowment and economic background? Is an exemption for a needs-based scholarship still justified?

The Treasury proposal asserted that a scholarship exemption is "unfair to the ordinary taxpayer who must pay for education with earnings that are subject to tax."[39] It is unclear, however, how much of the money paid by the "ordinary taxpayer" for her own education has in fact been subject to tax. For example, the student using family funds may have received funds as nontaxable gifts. Thus, perhaps the scholarship student could be viewed as receiving the nontaxable beneficence of a greatly "extended family". Both are using funds that, at least in their hands, were not subject to taxation. An exemption merely means that neither the gift recipient nor the scholarship recipient personally paid taxes on the funds used to purchase her education.

---

[38] For sample formulas, see 20 U.S.C. § 1087vv(a) (1988) (campus-administered grants) and 20 U.S.C. §§ 1070a-2(d), 1070a-3(c), 1070a-4(c) (1988) (direct federal grants).
[39] EXPLANATION OF PROPOSALS, *supra* note 6, at 62.

Unfortunately, this analysis is incomplete. Family gifts were probably subject to income tax in the hands of the relative who earned them. The scholarship recipient, on the other hand, is receiving funds that probably have not been taxed at all or that have resulted in a charitable deduction to someone. Both the family-contributed dollars and the dollars earned by the student herself will, therefore, be subject to tax, while the scholarship dollars will not. From this perspective, the Treasury's criticism of the exemption is justified.

A proper consideration of the Treasury's criticism, however, requires a consistent approach to transfers as well as a coherent view of the measuring unit for income. Yet the income tax currently adopts no coherent approach to either subject. Although all gifts and many transfer payments are excluded from income, there is no consensus about the appropriateness of this treatment. Some would argue that income should be measured in terms of total social product.[40] From this perspective, the scholarship recipient is being allowed to use pre-tax dollars while the family-funded student is using after-tax dollars. Many commentators feel that this disparity cannot be justified.[41] Some argue instead that income should be measured solely in terms of the recipient's well-being, without regard to the source of the funds involved.[42] From this perspective, both the family-funded student and the scholarship student are unjustifiably given benefits not available to the self-funded student. Still others argue that, although income should be measured from the standpoint of the taxpayer's well-being, this idea has to be tempered by a flexible notion of the taxable unit, one that would include at least the immediate family.[43] In this view, the self-funded or family-funded student is shortchanged vis-a-vis the scholarship-funded student.

Complicating this analysis still further is the fact that existing law shows no clear commitment to a notion of the appropriate taxable unit.[44] Under some provisions, the Tax Code adopts an expansive view of the taxpayer. For example, the unearned

---

[40] *See, e.g.*, Warren, *supra* note 15, at 1083–90.

[41] *See* sources cited at *supra* note 15.

[42] *See generally* Steuerle, *The Tax Treatment of Households of Different Size*, in TAXING THE FAMILY 82–83 (R. Penner ed. 1983).

[43] *See, e.g.*, McIntyre & Oldman, *Taxation of the Family in a Comprehensive and Simplified Income Tax*, 90 HARV. L. REV. 1573 (1977) [hereinafter *Taxation of the Family*].

[44] *See generally id.* at 1599–1602.

income of a child under fourteen is taxed as if it were the income of her parent, regardless of the source of the wealth from which that income is derived.[45] In other sections, however, the Code accepts the individual as the appropriate unit, regardless of living arrangements, for any other earnings of a household.[46]

In applying these notions to the student, the income tax recognizes the income earned by the college age student as the student's own income, and taxes it as such. However, to the extent she consumes and expends family funds other than her own, the expenditures are treated as the consumption of her parents. Without a clearer notion of the taxable unit, neither a justification nor a condemnation of a scholarship exemption solely in the terms used in the Treasury Department's criticism is possible.

The plight of the student who earns all of her own tuition costs remains problematic no matter how one resolves these issues. Even this student, however, is not treated unfairly as long as the needs-test for scholarship distribution accurately takes the student's ability to earn into account.[47] To the extent that the test cannot be precise, the exemption for scholarships merely exaggerates, rather than creates, the resulting inequity.

---

[45] I.R.C. § 1(i) (West Supp. 1990), added by § 1411(a) of the 1986 Act, now requires that the unearned income of a child under 14 be taxed at the same rate as her parent's income is taxed. Section 1(i)(7), added by § 6006(a) of the Technical and Miscellaneous Revenue Act of 1988, Pub. L. No. 100-647, 102 Stat. 3342 (TAMRA), in certain circumstances allows this income to be reported on the parent's return.

[46] The dependency definition and the exemption that relies upon it embody still another notion of the appropriate unit. Anyone who is a member of the taxpayer's household and lives at the taxpayer's home can be a dependent, I.R.C. § 152, but the exemption is available only for those dependents earning less than $2,000. I.R.C. § 151(c)(1). A child can remain a dependent even if no longer living in the taxpayer's household if (1) her income is less than $2,000, (2) she is under 19, or (3) she is under 24 and a full-time student. Prior to changes to § 151(c)(1), made by § 6010 of TAMRA, the continued exemption for a full-time student without regard to income was available regardless of the student's age. Prior to changes made to § 151(d)(2) by § 103(a) of the 1986 Act, the availability of an exemption on the parent's return did not prevent the child from also taking an exemption.

[47] This disparity in treatment between the earner and the transfer payment recipient is present whenever a transfer payment may be excluded. For example, the recipient of food stamps, public housing assistance, or any other kind of transfer payment is not taxed on the transfer payment, but the individual who earns only enough to provide herself with equivalent goods must bear the tax burden. The inequities of this situation have led at least one commentator to conclude that it is better to include even needs-based transfer payments in income and hope that exemption levels are high enough to limit the effect of the tax on the truly needy. *See* Pechman, *What Would a Comprehensive Individual Income Tax Yield?*, in 1 TAX REVISION COMPENDIUM 260 (1959). Even in this case there would be a discontinuity, because the federal government in effect would provide a matching grant to the scholarship student through the tax system and no matching grant to the nonscholarship student.

3. The Least Well-Endowed

The justification for the scholarship exemption on endowment grounds depends substantially on the means of scholarship distribution and its relationship to the distribution of cultural endowment. The more likely it is that the scholarships are distributed on a needs-basis and are available in a meaningful way to the least fortunate, the more sense it makes to treat them as substitutes for endowment that should not be subject to tax.

If this logic were pursued rigorously, however, one might conclude that the only scholarship system worthy of exclusion from the income tax is one which in fact tried to replace all "missing" endowment. This logic could lead to the conclusion that the only appropriate scholarship system is one that serves all the needs of those most handicapped—even if their handicaps make them least able to benefit from an education—before any of the needs of those more privileged are served.

Unfortunately, no matter how strictly needs-based the overall system of scholarship distribution is, it is not likely to reach those with the least initial endowment. Those with the least initial endowment are generally not in a position to pursue higher education at all. Thus, they will not receive the benefits of a scholarship program, much less a federal tax benefit for the scholarship program.

Perhaps, then, the tax law should take an even more aggressive role in equalizing endowments. Assuming that the income tax is to be used toward such ends, it should not just assist other redistribution programs. If equalizing endowments is a legitimate goal of the tax system, the tax law itself should respond completely by providing substantial subsidies to those who can demonstrate a lack of endowment.[48]

This amounts to an argument that an income tax is theoretically unsound unless it includes a "negative tax," or a system of payments to those whose endowment is below a specified level. Just as the income tax can be implemented without a negative tax, there is no reason for insisting that the income tax can only take an all or nothing role in endowment equalization.

---

[48] The earned income credit for low income taxpayers with children in § 32 could be thought of as this sort of endowment replacement. Like the scholarship provision, however, it is not available to those most lacking in endowment. Furthermore, the effect of the provision is more appropriately thought of as replacing the income from the missing endowment, rather than replacing the endowment itself.

The ultimate fairness of an intermediate position depends upon the other institutions to which endowment equalization is entrusted.

C. *Education as a Nonconsumptive Event*

1. Education as a Social Investment

The argument that all scholarships should be taxed relies upon the notion that the scholarship is paying the student's expense. But perhaps higher education is not the student's expense at all; it may instead be society's expense.[49] This investment in education is made by society as a whole to provide a better-informed citizenry, and thus, a better-functioning democracy, as well as to develop a more intelligent and responsive work force, thus improving overall productivity. This investment will also indirectly increase the total stock of basic and applied knowledge, further increasing productivity. Although this idea has been presented primarily in support of low tuition education and subsidies for educational institutions,[50] it has implications here as well.[51] A society's total investment in education should ideally provide an education for all without cost and without inclusion of the benefit in taxable income. The fact that we have insufficient means to meet this goal should not mean that those who do benefit from our incomplete attempts should be taxed.

Under this logic, those fortunate enough to receive tuition scholarships should not be subject to taxation any more than those who incidentally benefit from other social programs should be taxed. Scholarship recipients, for instance, should not be subject to taxation any more than those who happen to own frontage on a river that will be cleaned up as the result of new governmental regulation of upstream polluters. Although both

---

[49] *See, e.g.*, FINN, *supra* note 14, at 46. Two authors have concluded that the Massachusetts Bay Colony's support for a college that trained clergy to meet the community's spiritual needs was a social investment that triggered the first public subsidies to education. L. LESLIE & P. BRINKMAN, THE ECONOMIC VALUE OF HIGHER EDUCATION 28 (1988) [hereinafter LESLIE].

[50] *See, e.g.*, THE CARNEGIE COMMISSION ON HIGHER EDUCATION, HIGHER EDUCATION: WHO PAYS? WHO BENEFITS? WHO SHOULD PAY? 72 *passim* (1973). For a summary of the debate regarding whether this subsidy is better provided by student-directed or institution-directed aid, see EXPLORING THE CASE FOR LOW TUITION IN PUBLIC HIGHER EDUCATION (K. Young ed. 1974).

[51] *See, e.g.*, Bittker, *supra* note 23, at 938 n.23.

of these individuals benefit in a specific and quantifiable way—and perhaps even in a way detrimental to the interests of others—the incipient windfall should not be subject to the income tax.

An income tax will eventually be paid on the benefit as it is reduced to other forms traditionally thought to be more appropriate subjects of income taxation. The student will be taxed if and when greater incomes are realized in the future; there is no need to tax the student as she receives the education. The riverfront owner will be taxed when she opens a campground or sells her land for vacation homes and receives a greater return on her investment as a result of the clean frontage.[52]

Indeed, from this point of view, there is neither a need nor a justification for taxing the scholarship. Taxing the scholarship before the student is able to enjoy any real economic advantage is somewhat like taxing a new job based on the value that the new job holder is likely to be able to extract from the job in the future.

Several aspects of the way scholarships are in fact distributed buttress this argument that scholarships should not be taxed upon receipt. First, very few scholarships are granted in cash available to the student to spend as she pleases. Much is distributed only as credits against specific bills over which the student has little or no control beyond the initial choice of institution. The receipt of funds that cannot be spent as the recipient chooses should not be viewed as income to the same extent as receipts that can be spent as the recipient chooses. Second, most scholarships that are not distributed on a needs-basis are distributed on an ability-basis. Ability-based distribu-

---

[52] Some readers may see the analogy in the text as merely involving a realization question that is not present in the scholarship situation. Although the problem is in fact a realization question, the realization concept is also relevant in the scholarship case. In the case of scholarships, the realization question is harder to see because dollar values ordinarily can be assigned to the aid even if money does not change hands. But these dollar values, like those spent by the government for clean-up, are presumably buying a result for the donor; there is no reason, at least at the time the scholarship is given, to presume that the student is deriving any particular amount of value.

The relationship between the two cases may also be clarified if one looks at them from a consumption tax perspective. Both involve investments being made by third parties that incidentally benefit the taxpayer. Should an attempt be made at the time of the investment to identify the part of the investment that will ultimately redound to the taxpayer's advantage, or is it appropriate to wait until the investment is realized? Under the consumption tax ideal, it would be more appropriate to wait. For further consideration of the relationship between realization and the choice between income and consumption in defining the tax base, see Crane, *supra* note 27, at 199–200.

tions are consistent with the idea that scholarships involve a social investment made in a way that promises the greatest payoff to society as a whole, and the individual benefit is incidental.[53]

2. Education as Individual Investment

Even if scholarships are not best viewed as social investments, but rather as subsidies provided for a highly personal choice, it may not be appropriate to view their use as consumption. Educational costs represent investments in capital that, for many, are more like business investments than personal investments.[54] For such expenditures, some kind of cost recovery would be appropriate[55] but is denied under current law. Two

---

[53] Economists generally would not agree with the view suggested in the text that the individual benefit is separate from the social benefit. The more traditional approach is to view the individual benefit as part of the social benefit, or even as the best measure of the social benefit, and to view the social return as different from the private return only to the extent additional costs are included. *See* LESLIE, *supra* note 49, at 75–80. This view may be appropriate when looking at the impact of subsidized education over a lifetime or longer; it is less appropriate when attempting, as the income tax does, to fix values at particular, arbitrary points in time.

[54] Some researchers have questioned whether this view of higher education is rational. *See, e.g.*, Freeman, *Overinvestment in College Training*, 10 J. HUM. RESOURCES 287 (1975); C. BIRD, THE CASE AGAINST COLLEGE 116–35 (1975). Other studies have indicated that the private rate of return on college educations is still clearly positive, without counting either consumption benefits during the college years or psychic benefits of longer duration. *See* LESLIE, *supra* note 49, at 38–68.

[55] *See generally* D. BRADFORD, UNTANGLING THE INCOME TAX 206 (1986) [hereinafter UNTANGLING THE INCOME TAX]; R. GOODE, THE INDIVIDUAL INCOME TAX 80–92 (1976); Argrett, *Tax Treatment of Higher Education Expenditures: An Unfair Investment Disincentive*, 41 SYRACUSE L. REV. 621 (1990); Heckerling, *The Federal Taxation of Legal Education: Past, Present, and Proposed*, 27 OHIO ST. L.J. 117 (1966); McNulty, *Tax Policy and Tuition Credit Legislation: Federal Income Tax Allowances for Personal Costs of Higher Education*, 61 CALIF. L. REV. 1 (1973); Wolfman, *The Cost of Education and the Federal Income Tax*, 42 F.R.D. 535, 541–49 (1966); Comment, *Tax Treatment of Education Expenses: Perspectives from Normative Theory*, 55 U. CHI. L. REV. 916 (1988); *but see* BRADFORD, *supra* note 8, at 51; Brannon, *Scholarships, Loans and Tuition Tax Credit or Deduction in Taxation and Education*, in STUDENT AID IN HIGHER EDUCATION 135 (1966) [hereinafter Brannon] (arguing that the best argument for deduction is purity in defining the tax base, but questioning this on the grounds that education is already heavily subsidized, and that the net effect of a further reduction in college costs—"[a] higher tax burden on noncollege graduates for the benefit of college graduates making money"—seems inappropriate).

It is frequently suggested that if some sort of cost recovery is allowed for educational costs, the deduction should be disallowed if the funds used were exempt as scholarships on receipt. *See, e.g.*, GOODE, *supra*, at 84; Wolfman, *supra*, at 549; *cf.* McNulty, *supra*, at 27 n.85 (comparing family gifts with scholarships, especially if one views the scholarship as promoting the general welfare, rather than the student's welfare). If a deduction were denied for the use of scholarship funds, the exemption would in effect be eliminated. All education would in effect be paid for with before-tax funds, with scholarship funds being treated on a par with all other funds. An exemption should preclude a reasons are generally given for denying such recovery. First, there may be a substantial entertainment or luxury element to the expense. If this is true, the expense is clearly consumption and should be includible in income. Second, the appropriate rate of recovery for the investment in education is difficult to establish.

The exclusion for scholarships in effect ignores both of these problems with cost recovery. The exclusion of a scholarship receipt has the same effect as an immediate deduction for its use. Can this immediate deduction be justified? In some situations, an immediate deduction that is admittedly too fast might be preferred over some other form of recovery that is clearly too slow.[56] Such might arguably be the case for all education costs. If so, is there any justification for allowing a deduction for education costs only when they are met by tuition scholarship receipts, and not when they are met by personal and family funds?[57]

One ground for this distinction might be that educational expenditures funded by scholarships contain more of this "investment toward the future income" component than do other educational expenditures. Scholarship recipients, on the whole, may receive more benefits directly translatable into future streams of taxable income than nonrecipients, especially under a needs-based system. Further education, this argument would run, is more likely to increase the prospective earnings of the needy student than the affluent student. If this is true, perhaps scholarships should be treated differently than other educational expenditures, and some sort of cost recovery, or its surrogate, an exclusion, should be allowed. It is unlikely, however, that a student, because she receives a scholarship, actually makes more of an investment in her future than a student who enjoys

---

deduction for payments made with excluded funds only under the theory that the exclusion is in fact a substitute for the deduction. *See generally* Crane, *supra* note 27, at 262–66.

[56] The deduction under I.R.C. § 174 for research and development may be a good example. The deduction for research and development is treated as a current expense which is deducted in the year the expenditure is made, rather than as a capital outlay which would be expensed over a number of years. Such treatment provides an incentive for businesses to spend money on research and development.

[57] There is an alternative explanation for the fact that current law excludes tuition scholarships but denies a deduction for tuition paid. Under most older views of the income tax base, scholarship income, if income at all, is the income of the student, not her parents. Most of the political pressure for deductible education costs, however, supports a deduction for the student's parents.

a similar experience paid for out of her own personal or family funds.

A more likely ground for the distinction between the income tax treatment of these two students lies in the fact that almost all scholarship support is conditioned upon actually attending an institution of higher learning. The receipt of such restricted funds may not be worth as much to the recipient as the receipt of unrestricted funds, that is, funds that the recipient may spend as she pleases.

The unique success of price discrimination in the higher education market suggests that this doubt about the value of the scholarship to the recipient should not be overlooked. The total denial of a deduction for education costs amounts to a presumption that education is all luxury or entertainment, and therefore, is properly treated as an individual's consumption choice. Denying a deduction based on this consumption decision may conceivably be appropriate for those who choose to spend their own after-tax dollars on education, at least if the choice is between total denial and total allowance of the deduction. Extending this presumption about the consumption aspect of education to the scholarship student may be inappropriate.[58]

Again, the fit is not perfect; many scholarships fund educational experiences that are a luxury and many educational expenses that lack any luxury element are funded without scholarships.[59] But if the choice is either to include all scholarships

---

[58] This argument is strongest for those who accept the scholarship without knowing that it will be taxed. Once a recipient knows that the scholarship will be taxed, the recipient must value the education acquired at least as much as the tax liability she will face as a result of the scholarship. Recognizing this fact, however, does not help to provide a formula for determining how much that tax should be.

[59] An exemption for scholarship receipts as a substitute for cost recovery for education, even in the absence of more generalized cost recovery, could be justified on other grounds as well. The conceptual difficulty in determining whose income should be offset by the costs perhaps provides a significant reason why no deduction is allowed for college-level education. Frequently, the parent's funds are used in a way that will eventually generate income on the student's return. A deduction on the student's return at the time of the expenditure would likely be worth relatively little. A deduction on the parent's return would be worth more, but would be subject to the criticism that the deduction would be worth the most to the party who is the least sympathetic candidate for a deduction. In this case, transforming the deduction into a credit might help overcome the arguments about effects on progressivity, but questions about the proper return on which to claim the credit would remain.

If scholarships are distributed on a needs-basis, and need is determined with reference to the family unit, an exemption for scholarships would substitute for a deduction for education costs only in circumstances when a deduction would be least problematic, that is, where neither the parent nor the student could have made use of a deduction. The use of an exemption instead of a deduction finesses questions about the proper return on which to allow the deduction.

in income or to include none, including all may be the greater error.[60]

II. Scholarship Exemption as a Desirable Preference

Even if one rejects the above arguments for excluding educational assistance from the tax base on the ground that it need not be viewed as income, a strong argument can be made that, as a preference, the exclusion is far less objectionable than most preferences.

A. *Redirection of Resources Toward Education*

Most people accept the idea that the government should act in those domains in which the market cannot operate properly. It is plausible that the market will fail to provide the appropriate level of support for the kinds of activities traditionally associated with institutions of higher education.[61] It is also reasonable to assume that each individual's own education decisions may not reflect the optimal activity level. The limited information available to those making decisions about their education as well as the human inclination to discount the future benefit guarantee that these decisions will be inaccurate.[62] Support for education, whether targeted toward the student or the institution, is necessary to ensure that such activities are conducted at appropriate levels.

Conceding that markets cannot provide the proper level of educational activity means only that some governmental support

---

[60] Others would agree. *See, e.g.*, Goode, *supra* note 55, at 90 ("it is good social policy to resolve doubts [about the nondeductible personal component of education costs] in favor of more liberal writeoffs").

[61] *See generally* Friedman, *The Role of Government in Education*, in Economics and the Public Interest 123 (R. Solo ed. 1955), *reprinted in* M. Friedman, Capitalism and Freedom 85–107 (1962). For a survey of the literature on the question of the appropriateness of a subsidy for higher education, see generally, McNulty, *supra* note 55, at 42–57.

[62] Even if the individual accurately perceived the benefit from education, evidence suggests that many lower income individuals would not be able to adequately finance that education. Students would not be able to borrow through regular credit markets because of the relatively small principal amount involved, the lack of security ordinarily available given creditors' reluctance to view human capital as security, and students' lack of credit history. *See* Bosworth, *supra* note 14, at 130. While this market failure suggests that credit facilitation for students is appropriate, it does not necessarily imply that tuition subsidies are appropriate.

is necessary. This does not necessarily mean that the tax system should be invoked to provide the support. Tax preferences are rarely the most desirable form of governmental support. Most tax preferences are objectionable because they produce a disparity in the treatment of economic equals that would be unlikely to endure public scrutiny if provided in any other form. Exemptions for scholarships probably do not suffer from this infirmity to the same extent that other tax preferences do. For the most part, scholarships are distributed in ways that recognize economic inequalities, and at least in part, attempt to overcome those inequalities. Moreover, to the extent that the distribution of scholarships does not recognize economic inequities, it probably recognizes differences in ability, which is likely to translate into ability to create social benefit from the education. Public support for this kind of expenditure is likely to be relatively strong.

Tax expenditures may also be undesirable because, despite the apparent social goal served, they may produce undesirable economic consequences. That is, too much investment may be made in certain industries, and there may be too little control over the amount of benefit provided. These criticisms of tax expenditures have only limited application in considering the usefulness of an exemption for scholarships. The scholarship exemption merely makes other deliberate attempts to interfere with market mechanisms less costly. Unlike other more objectionable preferences, no taxpayer can enjoy a scholarship preference without having been identified by a third party using independent criteria as an appropriate recipient of a subsidy.

Using these standards, subsidies to education provided through scholarship tax exemptions are likely to be more efficient than those provided through some other means. With respect to federal scholarship subsidies, all that is lost in providing the exemption is accurate information about the total cost of the subsidy program. With respect to state and private scholarship subsidies, the additional cost is the same as that of any matching federal grant. In addition, under a scholarship system, the students themselves ultimately choose which institution to attend, and therefore, the one with which they will share the subsidy.

It is true that subsidies for higher education have come under considerable criticism. Some argue that institutions of higher learning have an insatiable appetite and will simply continue to spend whatever funds they are given without any improvement in the product they create.[63] There is also a growing skepticism with the argument that subsidies for higher education have in fact made education more widely available.[64] These problems are better left to those implementing the policies behind the scholarship distribution system generally. Tax policy, on the other hand, should not attempt to resolve these particular distributive concerns.

There are, however, problems with a tax expenditure for scholarships that should be treated as matters of tax policy, rather than general social policy. Tax policy must remain concerned about the preservation of the tax base and the integrity of income as an equitable measure of the capacity to pay tax. The exclusion of scholarships poses three potential threats to the tax base.

First, a general exclusion for scholarships poses a threat to the tax base when the scholarship ends up serving as payment for services on a long-term basis. Compensation for certain kinds of services could be systematically excluded from the tax base as a result of the exemption. This is unlikely to occur with respect to scholarship preferences for undergraduates. It can happen, however, either (1) when the educational process generates services for which third parties are willing to pay (as in the case of medical and veterinary training), or (2) when the process of learning and research becomes an individual's permanent livelihood.

The first problem can be contained by simply denying scholarship treatment when the educational process produces a marketable service or a service transferable to third parties.[65] No

---

[63] *See, e.g.*, Bennett, *Our Greedy Colleges*, N.Y. Times, Feb. 18, 1987, at A31, col. 1; Brannon, *supra* note 55, at 136 (suggesting that tuition tax credit would result in higher tuition charges). *But see* McPherson, Shapiro & Winston, *supra* note 30, at 32 (suggesting that in recent years, tuitions have tended to increase as federal aid has decreased, because, at least in the private sector, price discrimination is used to raise tuitions to fund aid).

[64] *See, e.g.*, Hansen, *Impact of Student Financial Aid on Access*, in THE CRISIS IN HIGHER EDUCATION, *supra* note 36, at 84, 91–96 (demonstrating that financial aid did not improve access, measured either in terms of enrollments or enrollment expectations, and suggesting as explanations that (1) programs missed the mark because of the extent they were watered down by political pressure to provide more aid to middle class, (2) aid was insufficient to replace the student's foregone income, especially with perceived lower incomes in the future, and (3) results could have been worse without the aid provided). *See generally* M. MCPHERSON, HOW CAN WE TELL IF FEDERAL STUDENT AID IS WORKING? (1988); LESLIE, *supra* note 49, at 134–80.

[65] This standard would probably resolve the more vexing questions regarding athletic scholarships. That is, when the expected athletic services are subject to sale by the

such limit was ever successfully introduced under prior law. The second problem goes to the heart of the justification for educational subsidies generally: should such subsidies be aimed at ensuring that a basic education is as widely available as possible or at furthering knowledge-seeking activities in general? This problem can be contained, as it is under current law, by limiting the exclusion to scholarships assisting students in degree programs. This limit ensures that the exemption is available only for educational programs of finite duration.

Second, a general exclusion for scholarships creates tension for the tax system when the scholarship provider receives services from the recipient. An educational institution may connect incidental tasks such as proctoring exams and cleaning bathrooms with financial aid. Providing exemptions for scholarships conditioned on such services is unlikely to shelter any individuals from income tax for any extended period of time because most of the students will not make a career of either activity. Nevertheless, allowing an exclusion for scholarships that in fact constitute payment for such services provides a subsidy to educational institutions not enjoyed by other tax exempt or for-profit entities.

If the intent is to provide a subsidy for education generally and to use the educational institution's criteria for establishing eligibility for that subsidy, extending tax-free treatment even to amounts that are clearly compensation paid by the educational institutions for services rendered to them might be appropriate. Indeed, linking payment to student willingness to provide services and become more involved in the institution might be desirable. However, the tax system must be able to limit this subsidy to the core educational activity intended to be benefited by the subsidy program. The problem lies in the difficulty of controlling this nature of the institution's activities.

Third, a general exclusion for scholarships also creates tension when an employer attempts to provide its employees with

---

school under terms similar to those that would be adopted by a commerical enterprise, no exemption would be available. Under current law, the Internal Revenue Service has been willing to afford athletic scholarships the same treatment afforded other scholarships so long as the scholarships are awarded "by the university primarily to aid the recipients in pursuing their studies." Rev. Rul. 77-263, 1977-2 C.B. 47. No reported case has been found, however, in which the Internal Revenue Service determined that this standard was not met. *See generally* Lee, *The Taxation of Athletic Scholarships: An Uneasy Tension Between Benevolence and Consistency*, 37 U. FLA. L. REV. 591 (1985); Randall, *Athletic Scholarships and Taxes: Or a Touchdown in Taxes*, 7 GONZ. L. REV. 297 (1972).

some degree of security regarding their children's education through scholarship programs from which the employees' children are more likely to benefit than the population at large. The inequities created by these exclusive scholarships may outweigh the considerations that justify the exclusion of more open scholarship programs.[66] Although control of this abuse is important, it should not be viewed as sufficient reason to eliminate all tax-free scholarships because it is not difficult to distinguish scholarships that are compensatory in nature from scholarships that are income only to the student, if income at all.[67]

B. *A Tax Expenditure for Education and Progressivity*

The most serious flaw of a scholarship exemption is a flaw of any exemption in a progressive tax system: any exemption or deduction is going to be worth more to a high-bracket taxpayer than to a low-bracket taxpayer, and will be worth nothing to the individual with no positive tax liability. At 1990 rates, an exemption for $10,000 for a taxpayer with $22,000 of other income will be worth about $2,800; it will be worth nothing to the taxpayer with only $3,000 of other income.

The importance of this problem again depends upon the methods used to distribute scholarships. If all scholarships were provided on a needs-basis that roughly paralleled the criteria used to establish income tax liabilities, the progressivity problem would be mitigated.[68] A tax expenditure in the form of an exemption will be more satisfactory when the scholarship recipient is in fact a low-bracket taxpayer.

Even if scholarships are not distributed solely on the basis of absolute need or need relative to the cost of the institution, the

---

[66] *See, e.g.*, Wheeler v. United States, 768 F.2d 1333 (Fed. Cir. 1985); Saunders v. Commissioner, 720 F.2d 871 (5th Cir. 1983); Armantrout v. Commissioner, 570 F.2d 210 (7th Cir. 1978).

[67] The problem of private foundations using their tax-sheltered funds to benefit related parties is a variation of this problem. The same standards that govern the circumstances under which charitable institutions that lack a broad base of public support can make gifts apply to this problem. A grant to an individual for study will result in a penalty tax on taxable expenditures under I.R.C. § 4945 unless the grant has an appropriate purpose, is an exempt prize, or is a scholarship "subject to the provisions of § 117(a) (as in effect on the day before the date of the enactment of the Tax Reform Act of 1986) . . . ." I.R.C. § 4945(g)(1) (1986), *amended by* Technical and Miscellaneous Corrections Act of 1988, Pub. L. No. 100-647, sec. 1001(d)(1)(B), 102 Stat. 3350 (1988). Furthermore, these grants must be made on an "objective and nondiscriminatory basis pursuant to a procedure approved in advance." I.R.C. § 4945(g) (1986).

[68] *See supra* notes 35–38 and accompanying text.

distributional effect of an exemption for scholarships may not be particularly troublesome. If the problem is simply that the scholarship system's definition of need does not comport with notions of equity embraced by the income tax, removing the exemption for scholarships is not likely to change greatly the ultimate distribution of benefit. A student's need will be determined with the subsequent tax burden in mind, and the amount of scholarship will be adjusted accordingly. Because the adjustment for the higher-bracket recipient will be greater than the adjustment for the lower-bracket taxpayer, removal of the tax exemption may actually enhance some measures of the regressivity of the total benefits provided.

More important, unlike many other kinds of tax preferences, the level of scholarship receipts is unlikely to be within the control of the recipient. Although one can make educational choices that can increase or decrease the amount of scholarship receipts, there are finite limits to an individual's ability to take advantage of the preference. One simply cannot use one's wealth, or the wealth of another through leverage, to take advantage of more of the tax preference for scholarships, as one arguably can do with the tax preference for other items such as municipal bonds or favorable cost recovery. It is therefore unlikely that the benefit of the scholarship exemption will gravitate exclusively to the highest brackets. Nor is it likely that another unpredictable group not already targeted by the substantive scholarship program will benefit disproportionately.

Finally, if the effect on progressivity is deemed sufficiently problematic, the traditional income exemption could be transformed into a credit whose value depended on income. This would ensure that the benefit is distributed proportionately to income level.

C. *Distortions Created by Arbitrary Definitions of Income*

Perhaps the most important problem with including scholarships in income is that failing to provide a tax expenditure for education through a scholarship exemption may substantially distort the way educational assistance is provided. As noted above, the tax base can never include all forms of educational assistance.[69] To the extent that the tax base includes only those

---

[69] *See* arguments presented in *supra* Section I.

parts of such assistance that are stated on tuition bills, and does not include those parts that are transferred without a written statement, educational institutions will feel pressure to provide benefits to all students without a specific charge even though some students may have been able and willing to pay separately. Furthermore, because education can be transferred in nontaxable ways in other contexts, notably in the employment context, the tax law would in effect create distortions in the kind of education that is pursued.

1. Grant Aid and Other Forms of Student Assistance

A distinction between values that can be transferred tax-free and those that are taxable could significantly impact the way educational assistance is provided, beyond those effects discussed above in the context of defining the tax base.[70] Although it is difficult to determine the degree to which decisions will be affected by such a tax difference, an income tax on educational assistance could prompt donors to provide aid directly to institutions, which could then transfer value to students in-kind and without tax, rather than to individual students. A state government, for instance, in deciding to provide funding for promising scientists, would be more likely to provide funds to institutions with science programs than to provide direct assistance to the student who could then pursue her education in the environment of her choice. Large donors in the private sector might respond similarly. Given the likelihood that this sort of aid distribution will not meet established federal goals as well as student-directed aid can,[71] these effects are undesirable.

In-kind subsidies would not be the only kind of educational assistance that would be favored if tuition grants were made taxable. A substantial portion of federal student aid is now given in the form of subsidized and guaranteed student loans.[72] Esti-

---

[70] *See supra* notes 28–30 and accompanying text.

[71] *See* sources cited *supra* note 30.

[72] The federal cost (that is, appropriated dollars) of student financial aid programs in fiscal year 1990 was approximately $4.8 billion for Pell grants and $3.8 billion in interest subsidy and payments on default. OFFICE OF STUDENT FINANCIAL ASSISTANCE, U. S. DEP'T OF EDUC., THE FEDERAL STUDENT FINANCIAL AID HANDBOOK 5 (1990–91) [hereinafter HANDBOOK]. These numbers say very little about what kind of aid is actually received: some aid is in the form of subsidized interest and placement fees, and some aid is in the form of payments on guarantees.

mates suggest that each $1,000 in principal amount loaned is equivalent to an immediate cash grant of $400 to $500.[73] The subsidy inherent in these implicit grants would be very hard to tax, because both the timing of the receipt and the identity of the recipient are obscured by the method through which the subsidy is paid.[74] Thus the form of federal aid most likely to be received by the middle class would remain untaxed, while the form of federal aid most likely to be received by those with lesser resources would be taxed.

Removing the tax burden from education aid that is directed at the individual will make that aid less costly, and will keep that form of aid competitive with other forms of aid to education. Properly focused, a subsidy provided through a tax expenditure can leave to the market at least some part of the allocation of education resources.

---

[73] BOSWORTH, *supra* note 14, at 135. This number covers only the cost to the federal government, based on certain assumptions about the federal cost of borrowing and its relationship to interest rates generally. The benefit to the student, for whom alternative means of borrowing might be available only at much higher cost or not at all, is much greater.

[74] The interest subsidy is unlikely to be included in the beneficiary's tax base. In the past, I.R.C. § 117 was cited as the sole authority for ignoring the interest subsidy. *See, e.g.*, Rev. Rul. 75-537, 1975-2 C.B. 32, *considered in* Gen. Couns. Mem. 33,721 (Jan. 4, 1968) (holding that interest paid pursuant to § 428(a)(1) of the Higher Education Act, Pub. L. No. 89-329, 79 Stat. 1219 (1965), by the Commissioner of Education on student loans is excludible as scholarship). The effect of the changes in § 117 on the logic of this memorandum is unclear. Because the entire amount is a payment of the interest costs of the student, rather than the tuition costs of the student, perhaps none of the interest subsidy should remain exempt. On the other hand, interest on principal amounts that were used to pay tuition, rather than room and board, may remain exempt. There is no indication from the government, however, that any effort will be made to include this interest subsidy in income.

In theory, any payment on the guarantee will be included as cancellation-of-indebtedness income. The Internal Revenue Service has taken a firm stand on the issue that loan forgiveness will result in taxable income, even though there are many circumstances in which the treatment of the forgiveness of a student loan will seem unduly harsh. The Service has indicated that the character of the benefit will be determined as of the time of the cancellation, rather than as of the time of the extension of the loan. *See, e.g.*, Priv. Ltr. Rul. 83-47-089 (Aug. 24, 1983) (indicating that when a student loan is cancelled, the time of the cancellation is the relevant time for determining the character of the forgiveness under I.R.C. § 108(f)); *cf.* Priv. Ltr. Rul. 87-14-035 (Jan. 2, 1987) (taxpayer who had borrowed money from employing educational institution and who was allowed to earn repayment credit had cancellation-of-indebtedness income; such credits were not scholarships because of the employment relationship, and I.R.C. § 127 was unavailable because "[w]hen the discharges took place the College was not providing educational assistance to [the taxpayer] to obtain [her] Master's Degree since that had already been completed," because the college was not "provid[ing] a specific incentive to [the taxpayer] to continue [her] education" and, because, in any event, there was no evidence of the plan required by § 127).

2. Tax Preferences for Liberal Education and for Workplace Education

Current law allows deductions only for education expenses that are clearly related to one's current income.[75] These standards allow education costs to be deducted only if the acquired education "[m]aintains or improves skills required by the individual [in her business]" or "[m]eets the express requirements of the [individual's job], imposed as a condition to the retention [of the job], but neither meets the minimal education requirements for a job nor qualifies the taxpayer for a new trade or business."[76] The regulations reflect concern that educational expenses may be primarily personal consumption, or that even if not personal, the expenses are capital expenditures with an indefinite useful life.

Current law also excludes from income any education provided in the course of one's employment. Prior to the enactment of the 1984 Tax Reform Act, the vague nature of such fringe benefits produced administrative results that can best be summarized as trusting the employer. If the employer thought that providing the education was appropriate, the Service would not second guess the employer and no amount would be included in income.[77] The 1984 legislation provided further guidance for

---

[75] Treas. Reg. §§ 1.262-1(b)(9) & 1.612-5 (as amended in 1967).

[76] Treas. Reg. § 1.612-5 (as amended in 1967). Courts have, for the most part, accepted these standards and applied them relatively strictly although the regulations themselves rather blatantly incorporate standards initially set forth in judicial opinions. *See, e.g.*, Walker v. Commissioner, 54 T.C.M. (CCH) 169 (1987) (conceding that although bar admission fees might be amortizable over the taxpayer's legal career, the cost of the bar review course could not be amortized). *See generally* Shaw, *Education as an Ordinary and Necessary Expense in Carrying on a Trade or Business*, 19 TAX L. REV. 1 (1963); Stephan, *Federal Income Taxation and Human Capital*, 70 VA. L. REV. 1357, 1407–13 (1984).

[77] The administrative position with respect to on-the-job training was articulated relatively late, probably because a taxpayer's victory in Bingler v. Johnson, 394 U.S. 741 (1969), would arguably have rendered all educational transfers in the employment context exempt under § 117. For a discussion of *Bingler*, see *infra* note 98 and accompanying text. There is still little authority regarding education provided by an employer at a jobsite when no payment that could be described as tuition is made to a third party. In Gen. Couns. Mem. 30,327 (Oct. 17, 1957), *cited in* Gen. Couns. Mem. 34,734 (Jan. 3, 1972), the Service apparently assumed that jobsite training would be exempt when it determined that a tuition program should not be included in gross income because it substituted for on-the-job training. Gen. Couns. Mem. 34,235 (Dec. 5, 1969) includes on-the-job training in its discussion of problematic facts, but does not discuss the problem further.

After *Bingler*, the Service appears to have taken the position that the standards of newly promulgated Treas. Reg. § 1.162-5 should apply to determine whether amounts paid as tuition to third parties by employers should be exempt for the employee. *See* the Service. Under section 132, a working condition fringe benefit is excludible if it would be currently deductible if paid for by the employee. For education expenses paid for or otherwise provided by an employer, this appears to mean passing through the hurdles of the regulations discussed above.[78] Much employer-provided training, whether on the job-site or provided by third parties, undoubtedly qualifies for the exclusion.[79]

These provisions ensure that education can be purchased in the employment context tax-free, either because an employee deducts what she pays or because the employer can provide it without generating taxable income for the employee.[80] Because

---

Gen. Couns. Mem. 34,235 (Dec. 5, 1969). This position was not made public, however, until 1976 in Rev. Rul. 76-71, 1976-1 C.B. 308, *considered in* Gen. Couns. Mem. 34,734 (Jan. 3, 1972). Shortly thereafter, I.R.C. § 127 was enacted, apparently in part to avoid the difficulties of applying the standards of the regulations in the context of courses taken under the direction of and paid for by employers. SENATE FINANCE COMM., GENERAL EXPLANATION OF THE REVENUE ACT OF 1978, at 125–28 (Comm. Print 1978); S. REP. No. 1263, 95th Cong., 2d Sess. 98 (1978).

[78] The assumption was repeated in the legislative history behind the 1990 extension of § 127, Statement of the Managers on Title XI of the Omnibus Budget Reconciliation Act of 1990, [1990] 47 Stand. Fed. Tax Rep. (CCH) 118 (extra ed. Oct. 29, 1990) outlined above have persisted. The legislative history of the Omnibus Budget Reconciliation Act of 1989, Pub. L. 100-203, 101 Stat. 1330 (1989), indicates that the standard, in the absence of the special provisions in I.R.C. § 127, would exempt employer-provided education only if it is "job-related," that is, if the education, "(1) maintains or improves skills required for the employee's current job, or (2) meets the express requirements of the individual's employer that are imposed as a condition of continuing employment." H.R. REP. No. 247, 101st Cong., 1st Sess. 900, 1171. This is essentially the standard articulated in the regulations under I.R.C. § 162, although it ignores the limitation regarding new skills contained in the regulations.

[79] *See, e.g.*, Hevener & Guarisco, *Fringe Benefit Regulations: New Details on Benefit Exclusions and Valuations Rules*, TAX NOTES, Nov. 6, 1989, at 743, 759.

For much on-the-job training, however, the tests in the regulations do not help at all. How should on-the-job training that is required for the job but not available outside the employment context be treated? Is this type of training necessary to keep one's job and therefore excludible, even if the training qualifies an individual for a new job and therefore would not be deductible to the employee herself?

Even if the equivalent of on-the-job training is available for a fee outside the employment context, it is not clear that the training should be taxed. Many employees, for instance, receive computer or accounting training, which could have been acquired by paying a third party. This training may be required by the employer, but it may also involve the acquisition of skills related to a new job.

The § 162 regulations were developed to protect the revenue base against the deduction not only of personal items, but also of capital items. The regulations reflect a decision that no deduction is preferable to a deduction that is taken too quickly, even when some deduction might theoretically be appropriate. The application of the regulations on the income side, however, will result in immediate and full inclusion with no later offset, even when such offset is appropriate. The rule for deductions that produced the smallest and most acceptable mistake may not produce a similarly small and acceptable mistake when applied to produce taxable income.

[80] Not only can the education be purchased tax-free within the employment context, but so can many of the incidental costs of acquiring that education. In most circumstances, if the cost of the education is deductible, so is the cost of whatever travel expenses (including transportation, room, and board) are incurred. *See, e.g.*, Johnson

such purchases are not allowed an equivalent tax benefit outside the employment context, the tax law perversely reinforces any market failure to support proper levels and types of education.[81] On the whole, only those individuals who are already employed and thus, probably least in need of the education, are given an advantage in purchasing education. Furthermore, this type of education is least likely to be undervalued in the market since it provides immediately marketable skills.

III. THE EXTENSION OF FAVORABLE TREATMENT FOR DIRECT EDUCATIONAL ASSISTANCE

Despite the callous consideration of scholarship exemptions in the legislative history of the 1986 Act, the 1986 changes can be understood as a compromise reflecting much of what is outlined earlier in this Article. The Act preserves the exemption for undergraduates to the extent that they receive assistance in paying for tuition, books, and various other education-related expenses; however, assistance for housing and other living costs is no longer exempt.[82] Like undergraduates, graduate students who are seeking degrees may exclude assistance for tuition, books, and other education-related expenses; however, graduate students who are not seeking degrees may not exclude such assistance.[83] How appropriate are these distinctions?

---

v. Commissioner, 55 T.C.M. (CCH) 700 (1988) (professor's travel costs from Texas to Hawaii in acquiring master's degree deductible).

[81] Because of its emphasis on existing employment, this scheme may have another perverse effect on the investments in human capital. If employers have more control over job definition than employees do, investments made in employment-related education will reflect the employer's sense of the value of various possible investments, rather than the employee's sense of such value. While a full examination of this effect is beyond the scope of this discussion, it is by no means clear that employer-directed education should be preferred over employee-chosen education.

The current income tax may prefer employer-directed education over employee-directed education in another, more subtle way. When employers supply the education, the employee's foregone income is not taxed; consequently, investments of foregone income in human capital are unaffected by the presence of an income tax. *Cf.* UNTANGLING THE INCOME TAX, *supra* note 55, at 204–06. In contrast, when the employee pays for the education in after tax dollars, the investment in human capital will not result in any cost recovery unless the requirements of Treas. Reg. § 1.162-5 (as amended in 1967) are met. As a consequence, the decisions of the potential student-investor will be biased in favor of taking low paying jobs that provide some educational opportunities on-the-job, rather than taking higher paying jobs and using the money earned to acquire education.

[82] I.R.C. § 117(b)(2).

[83] I.R.C. § 117(b)(1) & (2). *See generally* H.R. REP. No. 841, 99th Cong., 2d Sess. II-14 to -17 (1986).

A. *The Treatment of Assistance for Living Expenses*

Should the costs of meals and lodging incurred while pursuing an education necessarily be paid for from after-tax dollars? Behind this question lie difficult emotional issues concerning the value of living on campus as part of the educational experience and the socio-economic disparity between those who can expect the on-campus educational experience and those who cannot.

The Treasury Department's Report to the President simply offered the following explanation for distinguishing between tuition and other scholarship funds:

> In theory it might be appropriate to include the full amount . . . . In practice, this would create real hardships . . . . Scholarship awards are often made on the basis of need. If students were taxed . . . , they would often not have the resources to pay the tax. Moreover, . . . the recipient of a scholarship is not receiving an in-kind benefit in lieu of a cash amount and does not have the ability to convert the in-kind benefit to cash. The definition of income for tax purposes is appropriately limited by considerations of ability to pay. Accordingly, income from a scholarship for tax purposes should, in general be limited to amounts that represent out-of-pocket savings for regular living expenses.[84]

While the argument regarding the hardship imposed by the tax supports the exemption of all scholarships, the second argument, that funds received simply replace the ordinary cost of living, justifies the Code's distinction between scholarships for tuition and scholarships for room and board.[85] The differing tax treatment of the two types of scholarships seems appropriate under most approaches to the tax base, at least as long as one assumes that the student is not required to spend more on room and board while in school than she would otherwise spend. Even if the student spends substantially more on meals and lodging while she is in school, perhaps there is a point at which the

---

[84] EXPLANATION OF PROPOSALS, *supra* note 6, at 62–63.

[85] The House Report and the EXPLANATION OF ACT both argue that "the exclusion for scholarships should be targeted specifically for the purpose of educational benefits, and should not encompass other items which would otherwise constitute nondeductible personal expenses." H.R. REP. No. 426, 99th Cong., 1st Sess. 100 (1985); EXPLANATION OF ACT, *supra* note 5, at 40. The later reports did not attempt to justify the exclusion for scholarships generally, so it is impossible to determine whether the distinction between tuition and other scholarship aid is consistent with this definition. It is not clear, however, why *all* the non-tuition costs of education, for example, costs for transportation and room and board for those who would otherwise be living at home, are any more "personal" than the tuition costs.

student's choice of a college with the more expensive room and board must be seen primarily as a consumption choice of the student, rather than an education choice.

Many of the justifications offered above for an exemption for tuition scholarships are much weaker when considering an exemption for room and board scholarships. The subsidy enjoyed by students in public institutions is undoubtedly less for room and board than it is for tuition;[86] therefore, the problem of equating public and private pricing policies is far less severe in the context of room and board expenses. Since scholarship awards are likely to be greatest where education may be seen as the most extravagant, the argument that the least well-off must be served first is likely to be more compelling. Everyone must find a source of room and board regardless of whether she is in school; it is thus less appropriate to view room and board scholarships as a substitute for endowment. Similarly, the nondeductibility of the cost of room and board does not seem problematic.

The student, however, lacks control over the way her scholarship money is spent; she does not personally direct the funds that she receives, and she has little choice but to spend those funds in the way that they are directed. Because the student lacks control over her scholarship money, one may start to find a justification for excluding room and board from taxable income.

The conclusion that room and board scholarships should not be excludible is also questionable when compared with other situations in which a mixture of consumption and non-consumption features nevertheless results in treating the entire amount as non-consumption. For instance, the tax deduction for meals and entertainment is still available even when business is combined with significant amounts of pleasure.[87] However, the analogous argument regarding a deduction for room and board scholarships should not prevail. The better result would

---

[86] Indeed, it is probably nonsensical to try to establish any meaningful notion of the degree to which housing is "subsidized" at either public or private institutions. The housing market in the community is likely to be so dominated by the presence of the institution and its tax exemption that it is impossible to make useful statements about the institution's "costs" in providing room and board.

[87] *See generally* Halperin, *Business Deduction for Personal Living Expenses: A Uniform Approach to an Unsolved Problem*, 122 U. PA. L. REV. 859 (1974). *See generally* I.R.C. § 274 for guidelines for the minimum amount of business required to obtain a deduction.

be to treat all items the same—by treating as consumption any item that involves a substantial element of consumption.

Would a tax exemption for room and board scholarships be appropriate to implement existing federal attempts to aid an individual's educational pursuits? The answer depends upon the actual impact of the lack of exemption.[88] According to the history of the 1986 changes, the government assumed that, at current price levels, the inclusion of assistance for living costs in taxable income would rarely result in actual tax liabilities because most recipients of such aid would have low taxable incomes anyway after taking into account personal exemptions and the standard deduction.[89] This assumption may be true in

---

[88] As originally drafted, the limitation would have created insurmountable administrative problems. Section 123 of H.R. 3838 only allowed the exclusion for scholarships "to the extent the individual establishe[d] that, in accordance with the conditions of the grant, such amount was required to be used, and was used, for qualified tuition and related expenses." H.R. 3838, 99th Cong., 1st Sess. § 123 (1985). However, the committee report explained that the condition requirement meant that students did not have to trace their use of funds, provided the grant contained the appropriate conditions. H.R. REP. No. 426, 99th Cong., 1st Sess. 101 (1985).

Even with this liberal interpretation, however, many grants would not have qualified because a significant amount of grant aid is given based on need, without specification of the kind of cost reflected in that need. The statute as enacted provides that the exclusion is available only "to the extent the individual establishes that, in accordance with the conditions of the grant, such amount was used" for the allowed purposes. I.R.C. § 117(b)(1). The Conference Report clarifies that

> an otherwise qualified scholarship is not limited to a grant that by its express terms is required to be used for tuition and course-related expenses. Instead, the amount of an otherwise qualified scholarship . . . is excludable (taking into account the amount of any other grant to the individual eligible for exclusion) up to the aggregate amount incurred by the candidate for tuition and course-related expenses.

H.R. REP. No. 841, 99th Cong., 2d Sess. II-16 (1986).

The proposed regulations further reinforce the reality that actual tracing would not be required. Prop. Treas. Reg. § 1.117-6(e), 53 Fed. Reg. 21,688–701 (1988). Despite this legislative generosity, misinformation may have created problems for many scholarship recipients. At least one prominent guide to student aid asserts that separate accounts should be maintained to ensure compliance with the statute. J. MARGOLIN, FINANCING A COLLEGE EDUCATION 72–73 (1989).

[89] U.S. TREASURY DEP'T, TAX REFORM FOR FAIRNESS, SIMPLICITY AND ECONOMIC GROWTH—REPORT TO THE PRESIDENT 76 (Nov. 27, 1984) ("For most students, the higher tax threshold provided by the personal exemptions and zero-bracket amount will prevent the taxation of these benefits"); H.R. REP. No. 426, 99th Cong., 1st Sess. 100–01 (1985) ("In addition, under the bill, the committee has increased the income level at which individuals become subject to tax. Thus, grants of nonexcludable amounts based on financial need may not be subject to tax, if the amounts together with other income do not place the recipient above the taxable income threshold.")

For most dependent students, the 1986 Act replaced the zero-bracket amount of \$2,480 allowed in § 1 and an exemption of \$1,080 allowed by § 151 with a standard deduction of \$3,000 under § 63(e) and disallowance of a personal exemption. *See* discussion of the dependency exemption, *supra* note 46. Accordingly, the GENERAL EXPLANATION OF THE TAX REFORM ACT OF 1986 could not assert that higher thresholds were likely to reduce the possibility of any tax liabilities in fact resulting from the terms of absolute tax dollars, but it may not be true in terms of the student's budgeting problems. For instance, in 1986–87, the average room and board charge at public institutions was $2,720; the average at private institutions was $3,240.[90] From these numbers one can safely assume that substantial numbers of aid recipients at public colleges and at private colleges face a positive tax liability on their aid packages.

Even students receiving *only* federal aid based on criteria designed to assist the most needy could incur an income tax obligation at current price levels. Imagine a student attending her state's best public university with a tuition of $2,000 and room and board expenses of $3,000. She also incurs transportation costs, health insurance costs, and other costs includible in the federal definition of cost of attendance totalling $1,000. Further, suppose that this student's financial situation is such that she qualifies for the full Pell grant of $2,300 and an SEOG grant of $3,400.[91] This student may be a dependent for federal income tax purposes; therefore, she may not be entitled to a personal exemption and may get a standard deduction of only $3,000.[92] In such a situation $400 of her grants would be taxable

---

change. EXPLANATION OF ACT, *supra* note 5, at 40. For ease of exposition, the amounts assumed in the text are those provided for in the text of the 1986 legislation, without the inflation adjustments provided for in § 63(c)(4).

[90] EDUCATION STATISTICS, *supra* note 25, at 223. All of the statistics available at this writing report only average figures without providing information about the overall range of expenses. It seems safe to assume that if the average figures are approaching taxable levels, then substantial numbers have in fact exceeded it, at least for years before 1986. After 1986, such an assumption may be less appropriate, since institutions may have set their charged costs with a view to the level at which their assistance will become taxable.

[91] Federal aid to post-secondary students falls into three general categories: loans (both to students and to their parents); grants given directly to students, based on federal formula for determining need (Pell grants); and grants given by the educational institution itself according to more individualized formulas (Supplemental Educational Opportunity grants, state student incentive grants, work study). The overall pattern of aid resulting from these programs represents a somewhat unstable political compromise and has been subject to frequent revision in the last decade. *See generally* HANDBOOK, *supra* note 72, at 2–4.

[92] Most single taxpayers will not face a positive tax liability until they have income of more than $5,000 (a personal exemption of $2,000 and a standard deduction of $3,000); most married taxpayers will not have positive tax liabilities until they have $9,000 of income (two personal exemptions of $2,000 and a standard deduction of $5,000). A student's status as a dependent will, however, result in a lower tax threshold. Since 1987, under § 151(d)(2), if an exemption for a taxpayer can be claimed on another taxpayer's return, no other exemption for the taxpayer may be claimed. Therefore, full-time, single students under the age of 24 who receive more than half of their support from their parents will be taxed after $3,000. Scholarships are not included in the determination of support under § 152(d), so relatively small family contributions can result in dependency status.

Because a student who is claimed or is eligible to be claimed on her parent's return income. More surprisingly, out of every additional dollar of aid she receives, only $.85 will be available to pay the costs that justified the aid. Grossing up her aid will require that $1.17 be given for every additional dollar of identified need.[93]

A distinction between exempt tuition and fees and nonexempt living costs will undoubtedly lead to distortions in pricing policies of educational institutions. Educational institutions, particularly those providing substantial amounts of internal funds for student assistance, will feel strong pressure to change the way they account for the costs of housing and meals that they provide. A policy of providing much higher levels of subsidy for housing would be detrimental to the institution's ability to account accurately for its costs. In addition, this policy might detract resources from other subsidized programs.

On balance, the distinction drawn by the 1986 legislation between the tax treatment of tuition scholarships and room and board scholarships is a sound distinction. The irony in this conclusion, however, is that those students who receive aid above and beyond tuition aid are likely to be those with the least other resources available to pay the additional taxes incurred upon receipt of that aid.[94]

B. *The Treatment of Miscellaneous Costs*

The distinction the Tax Reform Act of 1986 draws between the tax treatment of scholarships for tuition and certain required expenses and the tax treatment of scholarships to meet a student's other needs will inevitably lead to changes in the way the educational product is packaged. First, more services will be provided without additional charge, yet the services actually will be paid for out of tuition and general revenue. For example,

---

will not be entitled to a personal exemption but can be treated as independent for financial aid purposes, either because she was not in fact claimed as a dependent, or because of an administrative decision to override the dependency presumption, persons entitled to low amounts of tax exemption may receive relatively large amounts of aid.

[93] Perhaps ironically, the 1990–91 edition of THE FEDERAL STUDENT FINANCIAL AID HANDBOOK reminds administrators to consider that college work study funds will be subject to applicable taxes but fails to mention the fact that income tax liability may reduce outright grant aid. HANDBOOK, *supra* note 72, at 7-5. For the student with no other taxable income, work study aid will be taxed at a marginal rate of 22.51%, while outright aid will be taxed at only 15%. Although includible in income, grant aid unrelated to services will not be subject to payroll taxes. *See infra* note 103.

[94] *See* Philipps & Bullivant, *The Ill Effects of Mid-1980s Tax Policy on Higher Education*, 6 AKRON TAX J. 45 (1989):

health services, recreational activities, computer access, and parking are more likely to be provided without charging additional fees. Second, more services that should be treated as discretionary expenditures are likely to be treated as required fees so that they fit under the tax exemption for tuition scholarships. Institutions with large internal scholarship programs are most likely to make these changes on a significant scale, since every dollar of tax avoided is an additional dollar of scholarship funds available for someone else.

C. *The Treatment of Payments Related to Services*

From its inception, the 1954 provision regarding scholarship payments related to services suffered from some ambiguity in its purpose.[95] Did the provision only make explicit the idea that scholarships were still seen as more analogous to exempt gifts than to other taxable awards, or did the provision intend to exempt scholarships despite the high likelihood of some sort of quid pro quo arrangement? Lurking behind this technical debate was a more fundamental question: should the provision be viewed as a general subsidy for academic pursuits, or does it reflect a more narrow intent to provide aid only for particular worthy recipients?

The provisions of the 1954 Code that were designed to resolve this question were hardly adequate. They provided that, for degree-seeking candidates, funds paid as compensation for services could not be excluded from taxable income unless all degree candidates were required to perform the same services.[96] Several reasonable interpretations of this language are possible. Congress may have intended to establish an irrebuttable presumption that payments were not for services if everyone had

---

[95] I.R.C. § 117 (1954).

[96] I.R.C. § 117(b)(1) (1954). Section 117(b)(1) provided that the exemption should not apply

> to that portion of any amount received which represents payment for teaching, research, or other services in the nature of part-time employment required as a condition to receiving the scholarship or fellowship grant.

The last sentence of that section provided that

> If teaching, research, or other services are required of all candidates (whether or not recipients of scholarships or fellowship grants) for a particular degree as a condition to receiving such degree, such teaching, research, or other services shall not be regarded as part-time employment within the meaning of this paragraph.

to perform the services to obtain the particular degree. Alternatively, Congress may have intended that compensation for services was nevertheless exempt except for degree candidates.

Section 117 has never recovered from its lack of clear purpose. If Congress merely intended section 117 to solidify the gift theory for the scholarship exclusion, then a factual inquiry into the possibility of a quid pro quo is clearly still intended. If, however, Congress intended section 117 to be a conscious extension of a subsidy for education, then the reasons for not exempting payments with a compensatory element are less obvious. Casual employment of the sort not likely to provide a career, such as library work or dishwashing, would not be troublesome, since the benefits to any one person would be finite. If Congress intended to subsidize all education, then scholarship status should be denied only in those cases in which the only activity an employee might ever do for the duration of a career looked plausibly like an educational activity. Even in the case of lifetime activities, scholarship treatment would not be inappropriate if Congress intended to subsidize all research activity of educational institutions.

From the outset, the Internal Revenue Service took a limited view of the purpose underlying section 117.[97] In *Bingler v. Johnson*,[98] the Supreme Court determined that no payment that was in fact compensation for services could be considered a scholarship. With its decision in *Bingler* the Supreme Court eliminated any chance for a liberal interpretation of section 117 based on the idea that educational institutions were worthy of special treatment.

Difficult questions remained after *Bingler*. One of the most frequently encountered problems has been presented by professional students, particularly medical and dental students, who undergo training after their formal coursework has ended. The heavily factual inquiry into scholarship status required under *Bingler* has given rise to an enormous amount of litigation in

---

[97] Under Treas. Reg. § 1.117-4(c) (1985), an amount was denied excluded status, even if paid "to enable [the recipient] to pursue studies or research, if such amount represents either compensation for past, present, or future employment services or represents payment for services which are subject to the direction or supervision of the grantor," or if paid "to enable him to pursue studies or research primarily for the benefit of the grantor." But, if the "primary purpose of the studies or research [was] to further the education and training of the recipient in his individual capacity," the grant was excludible.

[98] 394 U.S. 741 (1969).

such cases and has undoubtedly led to disparate treatment of taxpayers. Among the more vexing situations have been those presented by medical students, particularly those receiving support from institutions other than the institutions where their work was performed. Even if all recipients under a single program have been treated identically, their treatment may well have depended on the sophistication of the paying institution and of other outside sponsors in meeting criteria such as establishing the nature of the report to be completed or setting the terms of supervision, rather than on the actual merit of the situation.

Repeated litigation over this subject did not result in a coherent body of law.[99] Congress undoubtedly hoped that the litigation would disappear once it amended section 117 in 1986 to provide that no exemption is available to "that portion of any amount received which represents payment for teaching, research or other services by the student required as a condition for receiving" the scholarship.[100]

Unfortunately the 1986 provision changes very little, except to make clear Congress's displeasure with the amount of litigation on the question. No longer can any services be made an express condition of scholarships exempt under section 117. However, the statute sets forth no standard to apply when there is no such express condition and a student both works for and seeks scholarship aid from a single institution. What if research, particularly research that may be performed for hire, is required as a condition to obtaining a degree, and seeking a degree is a condition of the scholarship? Is the research then a condition of the scholarship? Similarly, what if an institution attempted to lower its costs by requiring all of its students to perform some sort of odd job on campus, such as working as a dormitory receptionist or dishwasher? Such a requirement should not ren-

---

[99] *See generally* Chommie, *Federal Income Taxation: Transactions in Aid of Education*, 58 DICK. L. REV. 93, 189, 291 (1954); Gordon, *Scholarship and Fellowship Grants as Income: A Search for Treasury Policy*, 1960 WASH U.L.Q. 144; Myers, *Tax Status of Scholarships and Fellowships*, 22 TAX LAW. 391 (1968); Stuart, *Tax Status of Scholarships and Fellowship Grants: Frustration of Legislative Purpose and Approaches to Obtain the Exclusion Granted by Congress*, 25 EMORY L.J. 357 (1976); Tabac, *Scholarships and Fellowship Grants: An Administrative Merry-Go-Round*, 46 TAXES 485 (1986); Tucker, *Federal Income Taxation of Scholarships and Fellowships: A Practical Analysis*, 8 IND. L. REV. 749 (1975); Note, *Federal Tax Incentives for Higher Education*, 76 HARV. L. REV. 369, 382 (1962); Comment, *Taxability of Scholarships and Fellowships*, 35 MO. L. REV. 393 (1970).

[100] I.R.C. § 117(c).

der all scholarship aid taxable.[101] It is unclear, however, whether a payment that is not a scholarship because of its relation to services will automatically be treated as wages.

Furthermore, in many instances, the line between merit and services can be a fine one, even at the undergraduate level. Should scholarships available only for those serving on the staff of a campus newspaper, perhaps sponsored by a sentimental alumnus, be considered compensation for services on the newspaper?[102]

Even if an amount of educational assistance is clearly taxable, determinations must still be made about whether services were provided, since scholarship grants are not subject to withholding or to payroll taxes.[103] Nor do scholarship grants

---

[101] Presumably the standard established for the threshold determination of scholarship is still good law. *See supra* note 97 (discussing the standard under Treas. Reg. § 1.117-4(c) (1985)). Under Treas. Reg. § 1.117-4(c), if the "primary purpose of the studies or research is to further the education and training of the recipient in his individual capacity," the grant is excludible. It is unclear whether the additional statutory language, denying exemption to any "portion of any amount received which represents payment for teaching, research, or other services by the student required as a condition for receiving the qualified scholarship," I.R.C. § 117(c), adds anything to this standard in cases not involving an express condition.

Prop. Treas. Reg. § 1.117-6(d) sets out the standard for determining when a scholarship will be considered to represent payment for services: if the relationship between the payor and the recipient requires that activities pursued are primarily for the benefit of the grantor, the payment will be considered a payment for services. Prop. Treas. Reg. § 1.117-6(d), 53 Fed. Reg. 21,688 (1988). This does not help distinguish the payor in its capacity as payor, seeking quid pro quo, from the payor in its capacity as educator, requiring from its students certain training activities, or the payor as substitute parent, requiring services to the community from all of its students, not just those receiving aid. Indeed, the regulations focus on the benefit received by the grantor, which deviates to some extent from existing case law. In Ferris v. Commissioner, 58 T.C.M. (CCH) 774 (1989), the student was required by the university to work for a local civic organization as a condition of receiving his work study aid. Even though the benefit to the payor from the services provided was tenuous, the court concluded that the payment did not meet the threshold test for scholarships.

The regulation provides further that the grantor is to determine what portion of the payment is for services, taking into account the amount that is paid to other nonscholarship students for similar services and the amount that is paid by other institutions for similar services. The regulation also makes clear that if a single sum is awarded, only that portion that represents payment for services is not excludible.

[102] *Cf.* Philipps & Bullivant, *supra* note 94, at 53 (noting this problem with respect to scholarships available for law review editors).

[103] I.R.S. Notice 87-31, 1987-1 C.B. 475, indicated the position of the Internal Revenue Service that amounts that would otherwise qualify as scholarships—but for the fact that they were spent on room, board, travel, clerical help, or equipment—would not be subject to withholding of income tax under I.R.C. § 3402, FICA taxes under § 3102, or unemployment taxes under § 3301. Furthermore, the notice made it clear that neither the grantor nor the educational institution through which the grant is administered need file a return of information with respect to the grant. This position regarding an information return was repeated in the proposed regulations. Prop. Treas. Reg. § 1.6041-3(q), 53 Fed. Reg. 21,688, 21,694 (June 9, 1988); *cf.* Rev. Rul. 71-378, 1971-2 C.B. 95 (amounts received by medical trainees excludible and such amounts not wages for employment tax purposes).

create an obligation on the part of the payor to report the payment.[104]

D. *Assistance for Graduate Work*

The 1986 Act allows exemption for tuition scholarships paid to all degree-seeking candidates at the undergraduate and graduate levels. Congress might not have been so generous, since other provisions providing special treatment for educational assistance, sections 117(d) and 127, do not provide an exemption for educational assistance at the graduate school level.

The arguments for excluding educational assistance on definitional grounds apply to some extent to graduate school education, at least if one sets aside the professional schools. The argument for excluding educational assistance based on endowment replacement is weakened by the fact that a graduate education is much more likely to move the student beyond a common baseline of endowment. However, the arguments based on the personal nature of the consumption probably still hold true.

The tone of the arguments must change, however, when considering assistance for professional education. Assistance for professional school education can be characterized less as an investment by society in the student than as an investment made on the student's behalf,[105] if only because the additional future income is so much more predictable. Professional school students are more likely to be spending tuition dollars in anticipation of future income; therefore, the lack of a deduction for their expenses is more likely to be erroneous. However, that same potential for future income makes an exemption for their educational assistance seem less equitable.[106]

If justified only as a preference, an exemption for graduate education poses a substantially greater threat to the revenue base than does an exemption for undergraduate education.

---

[104] *See* Rev. Rul. 75-537, 1975-2 C.B. 32 (student loan interest subsidy payments made by Commissioner of Education to lenders for interest not collected from borrowers under Higher Education Act of 1965 are scholarships under § 117(a) and excludible).

[105] This distinction is made sharply in Friedman, *supra* note 61.

[106] This seems clearly to have been Congress's motivation behind the limitations in I.R.C. §§ 117(d) & 127. *See, e.g.*, Report of the Commission on Ways and Means (Aug. 3, 1988) at 433 (discussing H.R. Rep. No. 795, 100th Cong., 2d Sess. (1988)) ("the exclusion does not apply to payment for . . . any graduate level courses of a kind normally taken by an individual pursuing a program leading to a law, business, medical, or similar advanced academic or professional degree").

Graduate students are more likely to be engaged in something similar to their life's work, and the institution is more likely to benefit from their services even if such services are not an explicit condition of financial assistance.[107]

E. *Assistance for Post-graduate Work*

The 1986 Act eliminated entirely the already limited exemption for fellowships for post-graduate work.[108] This sort of assistance was frequently for work that was closely related to the recipient's future occupation and could be justified only as an outright subsidy. Congress, perhaps overreacting to the apparent greed of medical students,[109] concluded that any such subsidy was inappropriate.

IV. The Exemption of Scholarship Assistance and Other Forms of Participant-Based Aid to Education

Finally, any discussion of an exemption for scholarships as a tax expenditure must take into account other tax preferences for educational assistance. Under virtually any criterion, except

---

[107] Congress appears to be unconcerned with this threat, since it has exempted tuition reductions provided to students working as research assistants and teachers, so long as they are otherwise adequately compensated for this work. *See infra* note 120.

[108] Prior to the 1986 changes, nondegree candidates could exclude scholarships and fellowships only if they were granted by an appropriate institution, and then only to the extent of $300 per month for a maximum of 36 months. I.R.C. § 117(b)(2) (1954).

Nondegree candidates could also exclude from their income any amount received to cover travel, research, clerical help, or equipment. I.R.C. § 117(a)(2) (1954). The statute now contains no special provision for such items because "in the case of grants to nondegree candidates for travel, research, etc., that would be deductible as ordinary and necessary business expenses, an exclusion for such expenses is not needed, and that an exclusion is not appropriate if the expense would not be deductible." EXPLANATION OF ACT, *supra* note 5, at 40. This explanation, however, overlooks the fact that some researchers will be receiving outside funds to pursue work they perform as employees, and therefore the deductibility of such expenses could be subjected to the two per cent limit in I.R.C. § 67.

[109] *See* EXPLANATION OF ACT, *supra* note 5, at 41 n.20 ("Why the amounts received by a young doctor just out of school should be treated differently from the amounts received by a young lawyer, engineer, or business school graduate has never been made clear" (quoting Zonderman v. Commissioner, 36 T.C.M. (CCH) 6, 9 (1977), *aff'd* 573 F.2d 1307 (4th Cir. 1978)).

perhaps a revenue criterion, the exemption for scholarships seems far more worthy of retention.[110]

A. *Employer Provided Educational Assistance*

An employer can provide up to $5,250 of educational assistance each calendar year to an employee if the assistance is provided under a nondiscriminatory program.[111] The assistance may be provided through tuition subsidies or through the actual provision of instruction; however, the assistance may not include payment for equipment, meals, lodging, or transportation.[112]

---

[110] It is sometimes argued that the largest single tax benefit granted education is the failure to tax the income foregone during the student years. S. Dresch, *College Enrollment*, in THE CRISIS IN HIGHER EDUCATION, *supra* note 36, at 112. Because this argument is based on a notion of a tax base inconsistent with most notions of a realistic income tax, it is ignored herein.

[111] I.R.C. § 127. Section 127 is one of a handful of tax expenditures subject to sunset. In the Omnibus Budget Reconciliation Act of 1989, Pub. L. No. 100-203, 101 Stat. 1330, § 127 was extended retroactively to include years beginning after December 31, 1988, and prospectively to include payments before September 30, 1990. Section 11403(a) of the Omnibus Reconciliation Act of 1990, Pub. L. No. 101-58, Subd. B, at 2 (1990) extended the provision for another year and closed the gap created by the fact that the provision had already expired for payments after September 30, 1990.

[112] The application of § 127 when the institution is the student's employer is unclear. The legislative history provides only hints. The Education Assistance Programs Act, Pub. L. No. 98-611, 98 Stat. 3176 (1984), added § 127(c)(8) to the Code, which allowed educational employers to provide tuition reductions under § 117(d) for graduate education to those among their employees who are research and teaching assistants. (This provision reinstated a provision that had been included in the House version of the Tax Reform Act of 1984 but had been deleted at Conference. *See* H.R. REP. No. 861, 98th Cong., 2d Sess. 1172 (1984)). Before this amendment, graduate education was covered under § 127 but not under § 117(d). It is unclear whether this provision was thought necessary because § 127 would not be available to the educational institution, or because the restrictions in that Code section were thought inappropriate.

Section 4001(b) of TAMRA temporarily made funding for graduate courses ineligible for purposes of § 127. It also added a new subsection, § 117(d)(5), to provide that graduate students working as teachers and research assistants are entitled to an exempt tuition reduction under § 117(d), subject to § 117(c). (Section 127(c)(8) seems to have been made redundant and was repealed by the Omnibus Budget Reconciliation Act of 1989.) The House bill contained amendments that implied that graduate students who worked as teachers and research assistants could receive support that was eligible under §§ 127 or 117(d), but the § 127 provision was dropped at Conference. Implicit in these provisions was the assumption that but for the new limitation on graduate education, § 127 benefits would be available to employees of educational institutions. H.R. 4333, 100th Cong., 2d Sess. § 301(b)(2)(A) (1988).

There are significant differences in benefit depending upon which provision controls. Section 117(d) still requires that the educational assistance not be compensation for services, while § 127 has no such limitation. The provision in § 117(d) is permanent, while § 127's is not. Section 117(d) is subject to antidiscrimination rules that are different from those to which § 127 is subject. Section 117(d) is not subject to the dollar limitation of § 127. Rev. Rul. 86-69, 1986-1 C.B. 78.

Under the current state of the law, the subsidy provided by section 127 for non-job related education is objectionable only in that it benefits those who already have jobs without providing an equivalent benefit for those without jobs.[113] Such a subsidy would be totally unjustified if education assistance paid outside the job context were subject to tax.

B. *Deductibility of Interest on Loans to Finance Education*

Prior to 1986, all interest on loans taken to support education was deductible.[114] Since 1986, interest paid on education loans is available only to those who have equity in a home that can be pledged as security at the time the loan is taken out.[115]

This provision, only incidentally thought of as a subsidy for education costs, is clearly a more problematic tax preference than almost any form of exemption for scholarships. It is available only to those who own equity in a home and benefits only those who otherwise face positive tax liabilities.

C. *Tuition Bonds*

Under TAMRA, interest on Series EE savings bonds is tax-free to the extent that, in the year of redemption, the taxpayer devotes the bond proceeds to her dependent's tuition costs after taking into account any other tax-exempt funds that might be available to meet those costs.[116] The exemption is phased out for taxpayers with high incomes.[117]

This special savings incentive also seems questionable in comparison to many kinds of scholarship assistance. It is available only to those who are able to accumulate capital before the educational costs must be paid and the provision provides no assistance for those who must pay for education after the costs are incurred.

---

[113] For plausible arguments justifying this position, see Stephan, *supra* note 76, at 1374.

[114] Under the 1954 Code, there were no general limits on deductions under § 163 for interest relating to personal indebtedness.

[115] I.R.C. § 163(h). Because interest on indebtedness is personal interest, interest on loans used to fund education will be deductible only if secured by the borrower's residence. *See* Priv. Ltr. Rul. 88-22-020 (Mar. 1, 1988).

[116] I.R.C. § 135.

[117] I.R.C. § 135(b)(2).

D. *Tuition Reduction*

Treas. Reg. § 1.117-3(a) originally treated the amount by which an educational institution "remitted" tuition charges for faculty children as a scholarship. This practice included employing institutions and cooperating institutions charging less than regular tuition and also employing institutions making payment to other institutions.[118] During the 1970's, this practice, as well as other fringe benefits, appropriately became the target of scrutiny.[119] In the Tax Reform Act of 1984, the situation was expressly addressed and the scope of permissible reductions narrowed. Section 117(d) now allows an educational institution to provide an employee or dependent of an employee a reduction in tuition "for education (below the graduate level),"[120]

---

[118] Treas. Reg. § 1.117-3(a) (1985) provides that

> [i]f an educational institution maintains or participates in a plan whereby the tuition of a child of a faculty member of such institution is remitted by any other participating educational institution attended by such child, the amount of the tuition so remitted shall be considered to be an amount received as a scholarship.

This regulation was based on a sentence contained in the Senate and House Reports that accompanied new § 117:

> If an educational institution . . . maintains or participates in a plan whereby the tuition of a child of a faculty member of any such institution is remitted at any other participating educational institution . . . attended by such child, the amount of tuition so remitted shall be considered to be an amount received as a scholarship under this section.

H.R. REP. No. 1337, 83d Cong., 2d Sess. 37 (1954).

Although the Service had not previously so limited its interpretation, in Knapp v. Commissioner, 867 F.2d 749 (2d Cir. 1989), the Service succeeded with its arguments that this language and the regulation derived from it applied only when the institution the student attended "remitted" the tuition by simply not charging the ordinarily required tuition, not when the parent's employing institution paid tuition to the child's school. *Cf.* Western Reserve Academy v. U.S., 619 F. Supp. 394 (1985), *aff'd*, 801 F.2d 250 (6th Cir. 1986) (cash tuition aid taxable to faculty members but not considered wages subject to withholding and FICA taxes for the institution providing assistance).

[119] Prop. Treas. Reg. §§ 1.117-3(a) & 1.117-4(c), 41 Fed. Reg. 48,132 (1976), *withdrawn*, 42 Fed. Reg. 3181 (Jan. 13, 1977).

[120] Section 127(c)(8), added by the Education Assistance Programs Act, Pub. L. No. 98-611, 98 Stat. 3176 (1984), provided that graduate students could enjoy a tuition reduction under § 117(d) if they worked as teachers or research assistants. Section 127(c)(8) was repealed by § 7814(a) of the Omnibus Budget Reconciliation Act of 1989, but its content was preserved in § 117(d)(5), added by § 4001(b)(2) of TAMRA.

Some tuition reduction may also be available under § 127 to students in graduate courses working in other capacities. *See* Priv. Ltr. Rul. 90-40-045 (July 10, 1990) (holding that § 117(d)(5) not available to full-time faculty and staff). This provision was made unavailable to graduate students by the amendment to § 127(c)(1) in § 4001(b)(1) of TAMRA, but restored by § 11403(b) of the Omnibus Budget Reconciliation Act of 1990, Pub. L. No. 101-58, Subd. B, at 2 (1990).

but only so long as certain anti-discrimination conditions are met.[121]

This provision seems to be aimed at the employee who is incidentally a student and whose assistance will be taxed under the main provisions of section 117, rather than the student who is incidentally an employee or is an employee only as a result of her need for financial aid. When it was first enacted, section 117(d) was intended to remove the taint of the employment relationship from tuition remission programs. The statute attempts to distinguish between those cases in which tuition remission is only an incidental part of compensation (*i.e.*, where the wages paid to those receiving and those not receiving such benefits are the same) and those cases in which tuition remission is the primary form of compensation. The amount of tension between these provisions and the new restrictions in section 117 concerning the person who receives the tuition

---

[121] Treas. Reg. § 1.132-1(f)(1) (1989) provides that

> [i]f the tax treatment of a particular fringe benefit is provided for in another section of Chapter 1 of the Internal Revenue Code of 1986, section 132 [does] not apply to such fringe benefit. . . . [B]ecause section 117(d) applies to tuition reductions, the exclusions under section 132 do not apply to free or discounted tuition provided to an employee by an organization operated by the employer, whether the tuition is for study at or below the graduate level.

The regulations do provide, however, that the working condition exclusion will still be available. Given the broad language of § 132(j) ("This section (other than subsection (e)) shall not apply to any fringe benefits of a type the tax treatment of which is expressly provided for in any other section of this chapter"), it is difficult to explain the availability of the working condition exclusion and not the employee discount exclusion.

One aspect of this overlap was addressed in § 7101(b) of the Omnibus Budget Reconciliation Act of 1989. It added § 132(h)(9), which provides that "[a]mounts which would be excludible from gross income under section 127 but for subsection (a)(2) thereof or the last sentence of subsection (c)(1) thereof shall be excluded from gross income under this section if (and only if) such amounts are a working condition fringe." As indicated by the title of the section, "Certain Otherwise Taxable Employer-Provided Educational Assistance May Be Excludible as Working Condition Fringe," (H.R. 3150, 101st Cong., 1st Sess. 7 (1989)) it appears that the drafters thought this language to be necessary to clarify the fact that failure to qualify under § 127 did not automatically result in inclusion in income, *see* H.R. REP. No. 247, 101st Cong., 1st Sess. 1171 (1989), even though the last sentence of Treas. Reg. § 1.132-1(f)(1) (1989) already so provided.

This sentence in the regulations and the new statutory provision run counter to the general rule in § 132(j) and Treas. Reg. § 1.132-1(f) (1989) that if a fringe benefit is covered by a specific Code section, the exclusions of § 132 are not available. Whether any negative implication is appropriate for educational assistance failing § 127 or § 117(d) for other reasons is unclear. Also unclear is whether payments for undergraduate education can qualify for exclusion under § 132 and, therefore, avoid the more restrictive provisions of § 127, including the payment limitation and the treatment of related meals, lodging, and transportation. A final unresolved issue is the effect of this statutory language on graduate tuition remission described in § 117(d), as amended by TAMRA, if such remission would qualify under § 127 but for the fact that graduate education is involved.

reduction and is also an employee remains to be seen,[122] as does the relationship between sections 117(d), 127, and 132.[123]

V. Conclusion

Given the difficulties inherent in defining the tax base, there is ample justification for excluding tuition scholarships. Tuition scholarships are only one of several kinds of educational subsidy that can be provided, and the other kinds of educational subsidies such as tax-subsidized state institutions and guaranteed loans are not susceptible to taxation. Tuition scholarships probably do not represent their full dollar's worth in consumption value, and there is no reason to believe that including the full dollar's worth will be a smaller mistake than excluding it entirely. Finally, given the rationale for providing preferences for educational assistance generally, the exemption for tuition scholarships seems unlikely to create undesirable distortions. Indeed, not providing an exemption for tuition scholarships may place those who seek education outside the workplace at a substantial disadvantage compared to those seeking education within the workplace. Such a disadvantage should be avoided both because those outside the workplace are likely to be more desirable beneficiaries of a tax preference for education and

---

[122] The General Explanation of the Tax Reform Act of 1986 purports to resolve this conundrum:

> The Act . . . explicitly provides that neither the section 117(a) exclusion nor the section 117(d) exclusion applies to any portion of the amount received that represents payment for teaching, research, or other services by the student required as a condition of receiving the scholarship or tuition reduction. If an amount representing reasonable compensation (whether paid in cash or as tuition reduction) for services performed by an employee is included in the employee's gross income and wages, then any additional amount of scholarship award or tuition reduction remains eligible for the section 117 exclusion as modified . . . .

Explanation of Act, *supra* note 5, at 43. In fact, however, the conundrum is no more adequately resolved for § 117(d) purposes than for § 117(a) purposes. *See supra* note 101.

[123] Ironically, the fact that § 117(d), and, probably, § 127, provide special rules for the kinds of educational assistance that educational institutions can provide for their employees precludes the application of the more general rules of § 132 relating to employee discounts. Thus, institutions providing graduate level education may therefore be among a handful of employers (who sell products specifically considered fringe benefits outside § 132) who cannot easily provide their services to employees generally at a discount. *See, e.g.*, Priv. Ltr. Rul. 90-040-045 (July 10, 1990).

because the kind of education sought outside the workplace is likely to be more in need of the subsidy.

The arguments supporting an exemption for scholarships to fund room and board are far less compelling. The arguments supporting exemptions for scholarships and fellowships beyond the undergraduate level may be as compelling, but concerns about erosion of the tax base and systematic inequity may lead to the conclusion that such exemptions are not appropriate.

STATUTE

LEASES OF PERSONAL PROPERTY: A PROJECT FOR CONSUMER PROTECTION

JOHN J.A. BURKE
JOHN M. CANNEL*

The leasing industry has grown so dramatically and creatively within the last decade that commentators, legislators, and judges have been unable to develop a consistent body of law that corrects the abuses and inequities of consumer-lease contracts. In this Article, Mr. Burke and Mr. Cannel propose a comprehensive statute to replace the ad hoc laws and doctrines presently governing consumer leases. In contrast to present leasing laws, the proposed Consumer-Lessee Protection Act fosters the uniform application of the Act by employing a reasonable and meaningful defintion of "consumer lease." Focusing on simplicity, certainty, and efficiency, the Act also eliminates the ambiquity and confusion presented by consumer-lease default formulas and risk allocations, and presents assignment of the lease as a viable alternative to default.

Consumer leases[1] have become a popular alternative to the secured transaction for financing the possession and use of per-

---

* John J.A. Burke and John M. Cannel are lawyers on the staff of the New Jersey Law Revision Commission. This Article is based on the work of this Commission, which is a public agency directed to revise the statutes of New Jersey and to consider recommendations from the National Conference of Commissioners on Uniform State Laws. N.J. STAT. ANN. § 1:12A-1–1:12A-9 (West Supp. 1990). In the process of considering Uniform Commercial Code article 2A for adoption in the state of New Jersey, the Commission drafted a consumer lessee protection act to cover consumer leases. N.J LAW REVISION COMM'N, REPORT AND RECOMMENDATIONS ON CONSUMER LEASES (1989).

[1] This Article adopts the definition of "lease" set forth in Uniform Commercial Code ("U.C.C.") article 2A which states that "a lease is created when the lessee agrees to furnish consideration for the right to the possession and use of goods over a specified period of time." U.C.C. § 2A-103(j) (1987). Whether a transaction constitutes a lease or security interest for purposes of article 2A is determined by the language contained in U.C.C. § 1-201(37). Several commentators question whether the distinction between leases and sales is valid. Ayer, *On the Vacuity of the Sale/Lease Distinction*, 68 IOWA L. REV. 667 (1983); Boss, *Leases and Sales: Ne'er or Where Shall the Twain Meet?*, 1983 ARIZ. ST. L.J. 357; Coogan, *Is There a Difference Between a Long Term Lease and an Installment Sale of Personal Property?*, 56 N.Y.U. L. REV. 1036 (1981). Others criticize article 2A for its failure to resolve this issue. Note, *Article 2A of the Uniform Commercial Code: An Unnecessary Perpetuation of the Lease-Sale Distinction*, 54 BROOKLYN L. REV. 1357 (1989); Comment, *Security Interests Under Article 2A: More Confusion in the Leasing Area*, 18 STETSON L. REV. 69 (1988).

This Article adopts an expansive definition of the term "consumer lease" based upon the model law for consumer lease transactions. *See infra* text accompanying notes 152–172.

sonal property.[2] There are three primary reasons for this development. First, consumer leases allow lessors to avoid compliance with the law of secured transactions. Second, leasing enables consumers to acquire goods they cannot afford to buy.[3] Third, lease contracts permit lessors to shift all risks of ownership to the consumer and retain essential rights of ownership. While leases serve the important commercial purpose of expanding access to the market, the technical language of the lease contract often conceals pitfalls for the consumer.[4] Although many of the pitfalls are now prohibited in secured transactions, they still exist in lease contracts because the leasing industry has outpaced the development of consumer leasing law.

Consumer leases present three primary problems that state and federal consumer protection statutes have failed to address adequately: (1) default and early termination penalties, (2) risk of loss and insurance costs, and (3) assignment of leases.[5] A more fundamental issue is raised by the different definitions of "consumer lease" found in current law and by the conceptual approach of current federal and state legislation. The absence of a uniform definition of "consumer lease" makes the class of beneficiaries entitled to statutory protection seem arbitrary and renders consumer leasing law confusing and incoherent. Furthermore, current federal and state consumer legislation embody the classical theory of economics, which has not functioned successfully to protect the consumer.[6] This theory of economics presupposes equality between consumers and merchants and ignores marketplace realities.

---

[2] In the rental industry, total revenues amounted to more than $13 billion in 1986. AMERICAN RENTAL ASSOCIATION, RENTAL INDUSTRY PROFILE (on file at the HARV. J. ON LEGIS.). The rental industry claims it does not lease goods because the rental period is short term. *Id.* The rental industry rents, for example, party supplies, medical and exercise equipment, light construction equipment, and do-it-yourself tools. *Id.*

[3] Leasing also has been spurred by the 1986 federal tax reforms which have phased out consumer loan interest deductions and removed the tax benefits of financing the purchase of goods. RENTAL INDUSTRY PROFILE, *supra* note 2.

[4] In addition, the fine print in lease contracts usually contains onerous terms for default and early termination of the lease. *See* Wall St. J., July 19, 1989, at B1, col. 3.

[5] Lease-purchase agreements, which are not the focus of this Article, raise an additional problem of cost. These agreements are better treated as sales subject to consumer credit legislation. The New Jersey Law Revision Commission (the "Commission"), therefore, recommended an amendment to the New Jersey Retail Installment Sales Act to include lease-purchase agreements within the definition of "retail installment contract." N.J. LAW REVISION COMM'N, REPORT AND RECOMMENDATIONS ON CONSUMER LEASES (1989), and appendix to this article at app. F § 3(a).

[6] The reference is to the free market economic theory discussed *infra* text accompanying notes 99–101.

The federal government and four states have enacted consumer protection legislation aimed at consumer leases.[7] The federal statute requires disclosure of key lease terms and contains a general prohibition against unfair and unreasonable commercial practices.[8] The latter provision does nothing more than codify the contract law principle of "unconscionability," a principle that has not been effective in providing adequate consumer protection in lease transactions. Massachusetts and Washington have statutes that simply reproduce with some modifications the provisions of the federal law.[9] California and Maryland have statutes which resemble consumer credit legislation and apply only to motor vehicle leases.[10] These two statutes, applying solely to consumer lease transactions, fail to regulate penalties upon default, insurance costs, and assignment in leases of personal property.[11]

Current statutes fail to protect lessees mainly because consumer protection law evolved to correct abuses in the sale of goods and did not anticipate the advent of leasing.[12] Consumer law focuses almost entirely upon the suppression of deceitful practices in the sale of goods, disclosure of information about contract terms, and the control of maximum finance rates in secured transactions.[13] This limitation has led courts to examine leases within the conceptual parameters of a sale, and to extend consumer protections only to leases that approximate a sale of goods.[14] Even if courts were to apply existing consumer protec-

---

[7] Louisiana has comprehensive code provisions governing leases of personal property. LA. REV. STAT. ANN. §§ 9:3301–9:3342 (West Supp. 1989); *see also* LA. CIV. CODE ANN. art. 2668–2744 (West 1952). These provisions are analogous to article 2A of the U.C.C. and do not deal with consumer issues.

[8] THE CONSUMER LEASING ACT, 15 U.S.C. § 1667 (1988).

[9] MASS. ANN. LAWS ch. 93, § 90 (Law. Co-op. Supp. 1990); WASH. REV. CODE ANN. § 63.10.010–63.10.900 (1990).

[10] CAL. CIV. CODE § 2985.7 (West 1974 & Supp. 1990); MD. COM. LAW CODE ANN. § 14-2001 (Supp. 1990).

[11] *See infra* text accompanying notes 33–37.

[12] D. ROTHSCHILD & D. CARROLL, CONSUMER PROTECTION: TEXT AND MATERIALS 4–18 (1973); Barber, *Government and the Consumer*, 64 MICH. L. REV. 1203, 1205–06 (1966); Miller, *Consumer Leases Under Uniform Commercial Code Article 2A*, 39 ALA. L. REV. 957–59 (1988).

[13] Barber, *supra* note 12, at 1215.

[14] *See, e.g.*, Sheffield Commercial Corp. v. Clemente, 792 F.2d 282 (2d Cir. 1986) (finding automobile lease constituted a sale subject to Motor Vehicle Retail Installment Sales Act); *In re* Tillery, 571 F.2d 1361 (5th Cir. 1978) (finding automobile lease constituted a security interest subject to U.C.C. article 9); Barco Auto Leasing Corp. v. House, 202 Conn. 106, 520 A.2d 162 (1987) (finding automobile lease constituted a sale subject to Retail Installment Sales Financing Act and Unfair Trade Practices Act); Keeling v. Ford Motor Credit Co., 314 Md. 311, 550 A.2d 932 (1988) (finding automobile lease did not constitute a sale governed by Retail Installment Sales Act); Ford Motor

tion laws to leases, these laws would not adequately protect lessees because existing consumer protection laws are an unsuitable means of ensuring fair lease contracts.[15]

In fact, article 2A of the U.C.C. exacerbates the shortcomings of existing consumer law.[16] Article 2A, intended as a comprehensive codification of the law with respect to leases of goods, established the rules for commercial transactions.[17] Currently enacted in nine states and pending in many others,[18] this U.C.C. article actually reduces consumer protection in lease contracts, because the drafters included only a few consumer protection provisions within the article itself. Moreover, most of these provisions can be excluded by contract. While the freedom of contract principle is appropriate for commercial transactions which presuppose parties of equal bargaining power, it is inappropriate when it subjects consumer leases to the whims of the commercial lessor. Although the provisions of article 2A are subject to the applicable consumer protection laws of the enacting jurisdiction,[19] no jurisdiction has yet to enact effective consumer protection laws for leases.

This Article proposes a model act to resolve the consumer issues left open by article 2A and not addressed by existing law.

Credit Co. v. Sims, 12 Kan. App. 2d 363, 743 P.2d 1012 (1987) (finding lease did not constitute sale subject to lemon law); Sellers v. Frank Griffin AMC Jeep, Inc., 526 So. 2d 147 (Fla. Dist. Ct. App. 1988) (finding automobile lease did not constitute sale covered by U.C.C. article 2 and the Magnuson-Moss Act).

[15] Most state unfair and deceptive acts and practices statutes apply to leases, but these statutes primarily regulate marketing devices. *See infra* text accompanying notes 66–69.

[16] Article 2A of the U.C.C. governs all leases of goods and applies to consumer leases. U.C.C. § 2A-102 (1987). The article contains nine consumer protection provisions: U.C.C. §§ 2A-106, 2A-108(1), 2A-108(2), 2A-108(4), 2A-221, 2A-309, 2A-406, 2A-504, 2A-516(3); *see also* Miller, *supra* note 12, at 966 (analyzing consumer provisions of article 2A). Consistent with the policy of the Code to defer local matters to state legislation, comprehensive consumer protection provisions were deliberately omitted from article 2A. *Id.* at 962. To the extent that article 2A contains few consumer protection provisions, its role as a consumer protection statute in lease transactions of personal property is almost insignificant.

[17] U.C.C. §§ 2A-101 to 2A-531 (1987). *See also Symposium: Article 2A of the Uniform Commercial Code*, 39 ALA. L. REV. 559 (1988) (reviewing article 2A).

[18] Article 2A has been enacted in the following states: California, CAL. COM. CODE § 10101 (West 1990); Florida, 1990 FLA. SESS. LAW SERV. Ch. 90-278 (West); Kentucky, 1990 KY. ACTS Ch. 63; Minnesota, MINN. STAT. ANN. § 336, 2A-101 (West & Supp. 1990); Nevada, NEV. REV. STAT. § 104A-2101 (Michie Supp. 1989); Oklahoma, OKLA. STAT. ANN. tit. 12A § 2A-101 (West & Supp. 1990); Oregon, 1990 Or. Laws ch 676, §§ 1–78 (Supp. 1990); South Dakota, S.D. CODIFIED LAWS ANN. § 57A-2A-101 (Supp. 1990), and Utah, UTAH CODE ANN. § 70A-2A-101 (Supp 1990). Many states are considering adoption of article 2A. NAT'L CONF. OF COMM'RS OF UNIF. STATE L, STATE LEGISLATIVE ACTIVITY REPORT (Oct. 1, 1989).

[19] U.C.C. § 2A-104(d) (1987). *See* Miller, *supra* note 12, at 963.

The statute, entitled the "Consumer-Lessee Protection Act" (the "Model Act"), is based upon an act proposed to the New Jersey legislature by the New Jersey Law Revision Commission.[20] The Model Act regulates the substantive content of consumer lease contracts in the three principal problem areas outlined above and allocates the risks and resources of property in this limited context. The Model Act also rejects the conceptual approach to market regulation taken by current consumer legislation, by directly intervening in the marketplace to control the business relationship between the consumer and lessor and establishing ethical standards of conduct. Finally, the Act broadens the class of beneficiaries of consumer protection. While the Act is meant to complement article 2A, it is designed to stand alone as a model consumer lease protection law.[21] As such, it accomplishes for consumer leasing law what article 2A accomplished for commercial leasing law.

Part I of this Article defines the issues in consumer lease transactions which existing law and article 2A fail to resolve, and demonstrates the failure of current approaches to regulate consumer lease contracts. Part II presents the provisions of the Consumer-Lessee Protection Act, a model statute which resolves the issues identified in Part I. Part III explains the policy reasons for broadening the definition of "consumer lease" and for rejecting the economic and contract theory embraced by current federal and state consumer legislation. The complete text of the Model Act is contained in the Appendix.

I. Issues

A. *Default and Early Termination Penalties*

The main problem with consumer lease contracts lies in the method used to measure damages upon default or early termi-

---

[20] The Model Act varies significantly from the Commission version. The definition of "consumer lease" is far broader in the Model Act than it is in the Commission version. *See* N.J. Law Revision Comm'n, *supra* note 5, at app. F § 3(a). The Model Act also contains two alternative risk of loss provisions, Alternative A and Alternative B, while the Commission version contains only Alternative B. *See id.* at app. F § 12. The default and insurance provisions of the Model Act also differ slightly from the Commission version. *See id.* at app. F §§ 13, 16. The Model Act is cited as Model Act with the appropriate section number. The Commission version of the Consumer Lessee Protection Act has been introduced in the New Jersey Senate. S. 2791, 204th Leg. Sess. §§ 1–21, N.J. 1990. The Act is also under study in Connecticut.

[21] *See supra* note 20.

nation of the lease. The combined penalties and damages imposed upon default and early termination of the lease can often exceed the amount the consumer would have paid under the lease had the consumer completed performance.[22] Thus, the consumer is charged more than is necessary to make the lessor whole. Moreover, consumers are not always informed of their potential liability because the formulas used to calculate default and early termination penalties, even when disclosed, are not always adequately explained.[23]

To understand the operation of default and early termination penalty provisions, one must examine the economic structure of consumer lease contracts.[24] In a consumer lease, the lessor purchases the property to be leased by the consumer and thus incurs initial financing costs. The lessor then rents the property to the consumer for a period of time less than the economic life of the goods specified in the lease. The expected economic value of the goods at the end of the lease term often is specified in the lease contract. The total financial obligation assumed by the lessee consists of discrete elements: (1) the initial cost of the property to the lessor, (2) the depreciation in the value of the goods over the period of the lease, and (3) the lessor's profit. The lessee pays the total amount due under the lease in equal monthly installments over the duration of the lease term. When the lessee makes all monthly payments due under the lease, the lessor receives its initial costs, compensation for depreciation

---

[22] This problem is best illustrated by closed-end auto leases. *See* NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES § 5.4.4 (2d ed. 1988); *see also* NAT'L CONSUMER L. CENTER, TRUTH IN LENDING app. O at 110 (Supp. 1988); Closed End Motor Vehicle Lease and Disclosure Agreement 1756-6, Item 13 (1989) [hereinafter Closed-End lease]; Maryland National Automobile Lease Agreement 783-14, Item 35 (1989) [hereinafter Maryland lease]; Lease and Go, Inc. Motor Vehicle Lease Agreement and Disclosure Statement Closed End 0447, Item 19 (1988); Nissan Motor Acceptance Corporation Lease 3001-B (clause H and I) (October 1987) [hereinafter Nissan lease]; General Motors Acceptance Corporation Lease 871 DLP (July 1985); Bavarian Motors Lease Agreement-Consumer Form CFD-265 CLA (Mar. 1977) (all leases referenced within this Article are on file at the HARV. J. ON LEGIS.).

[23] A lease used by Citicorp National Services, Inc., which computes the early termination charge by the Sum-of-the-Digits method (also called the "Rule of 78") does not define the latter term. Vehicle Lease and Disclosure Statement SF-7075, Item 17 (Jan. 1990) [hereinafter Citicorp lease]. The Sum-of-the-Digits method is a formula for calculating unearned interest on prepayment of an installment loan. Its use persists despite the possibility of more accurate methods of calculation because it favors the lender. The difference between the amount of interest rebated using the Rule of 78 and the actuarial method can be substantial, especially when a long term loan is prepaid. P. ROHNER, THE LAW OF TRUTH IN LENDING, para. 5.05[11][b] (1984).

[24] *See* NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES, *supra* note 22, §§ 5.4.4.2, 5.4.4.2.1 for a more thorough explanation of the economic structure of leases.

in the value of the goods, and the amount of profit anticipated at the beginning of the lease. This economic structure works well if the lease contract remains in force for its full term.

If the lessee defaults or terminates the lease contract early, however, the amount received in lease payments up to the point of default or early termination may not compensate the lessor for its costs and anticipated profit. This may occur because the lessor's total costs are spread unevenly over the term of the lease contract. The initial cost of the property is incurred at the beginning of the lease. The depreciation cost, however, is incurred at a variable rate over the term of the lease. Because the goods depreciate in value to the greatest extent during the early stages of the lease, the monthly payments received by the lessor do not equal the lessor's depreciation cost. Other costs, such as operating and capital costs, may also be incurred at a constant rate over the term of the lease. While the lessor is entitled to receive compensation upon an early termination or default of the lease contract, the formulas used in most default and early termination clauses produce depreciation deficiency amounts that far exceed the lessor's actual costs plus any reasonable profit.

The classic default formula charges the lessee the past-due lease payments and all future payments due for the duration of the lease.[25] To offset this amount, the lessee receives a credit for the difference between the estimated residual value of the goods specified in the lease itself and the value of the goods realized after disposition by the lessor. No other credits are given to the lessee. A standard automobile lease illustrates the unfair results of the classic default formula.[26] Consider a $10,000 car with a $4,000 residual value and a $250 monthly lease payment for four years. Assume the consumer defaults after four months, having made three lease payments, and the lessor sells the car at wholesale auction for $7,000. Under the classic default formula, the consumer is liable for the $250 delinquent lease

[25] The term "classic default formula" refers to formulas used to calculate damages upon default or early termination that do not reduce future payments to present value. NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES, *supra* note 22; *see also* R. ABRAMS, NEW YORK STATE ATTORNEY GENERAL'S REPORT: "EARLY TERMINATION" CLAUSES IN LONG TERM AUTOMOBILE LEASES: CONSUMERS ARE PAYING TOO MUCH (1989) (reviewing several industry leases to determine the effect of classic default).

[26] This example is taken from NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES, *supra* note 22, § 5.4.4.2.1.

payment and $11,000 in remaining lease payments (forty-four remaining months times $250). Against this liability of $11,250, the consumer is given a credit of $3,000 for the difference between $7,000 and $4,000. The claimed deficiency is thus $8,250. The lessor receives $16,000 ($750 in lease payments made prior to default, $7,000 resale of the car and a $8,250 deficiency) for a $10,000 car after four months. The lessor anticipated it would receive the same sum of $16,000 after four years.

The classic default formula gives the lessor a substantial economic benefit upon default or early termination without giving the lessee a credit for the early payment of the total lease obligation. The lessor receives $16,000 immediately when it was only entitled to receive this amount in equal payments over a four-year period. The economic effect of receiving the money early is substantial, amounting to a windfall between $2,500 and $3,500.[27] This is an inequitable result, especially in a consumer transaction. The lessor should not have the right to a full return upon default; when the leased car is returned, the lessor is no longer deprived of its initial investment.[28] The lessor should give the lessee a credit for the unearned lease charges.[29]

The classic default formula also does not give the lessee credit for the economic benefit of the early return of the goods.[30] The lessee receives a credit for the difference between the realized value and the estimated residual value of the goods which is specified in the lease. However, the residual value is what the lessor may claim at the end of the lease. The right to receive the residual value immediately is worth more than the right to receive the same amount at the end of the lease term because money received earlier will be increased by the interest it earns. Therefore, the credit to the lessee should be the difference

---

[27] The formula for the present value of a stream of future payments is:

$$\frac{P[1 - (1 + i)^{-n}]}{i}$$

where P is the amount of the monthly payment; i is the monthly interest rate; and n is the number of payments. Assuming an interest rate of 12% (1% per month), the present value of $16,000 due as 48 monthly payments of $333.33 is $12,657.99. Assuming an interest rate of 9% (3/4% per month), the value is $13,394.93.

[28] NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES, *supra* note 22.

[29] *Id.*; R. ABRAMS, *supra* note 25, at 6–7.

[30] NAT'L CONSUMER L. CENTER, UNFAIR DECEPTIVE ACTS AND PRACTICES, *supra* note 22; R. ABRAMS, *supra* note 25, at 9–10.

between the realized value and the present worth of obtaining the residual value.

The method used in the classic default formula to calculate the value of the property after default tends to establish a low value for the goods.[31] A lessor typically sells the property at public or private auction. This method of disposition often produces a sales price lower than the estimated residual value of the property specified in the lease contract. Consumers cannot be expected to bring prospective buyers to the auction to increase competition for the goods. Moreover, lessors cannot be expected to promote buyer interest in an auction where there is no incentive for them to do so. The actual result of a particular auction depends upon numerous variable factors and produces inconsistencies in the sales price. To the extent that the credit the consumer receives for the realized value of the goods does not equal the actual value of the property, the reduction serves to increase the deficiency claim the consumer must pay.

The importance of a fair sale to the consumer is demonstrated by the significant protections afforded the consumer in secured transactions. In secured transactions, the disposition of consumer goods after default is governed by article 9 of the U.C.C.[32] All aspects of the sale must be carried out in a commercially reasonable manner.[33] The creditor must give notice of the sale to the consumer,[34] and if the creditor disposes of the goods in a commercially unreasonable manner, or fails to give the required notices, the violation may bar the creditor's right to claim a deficiency.[35] Despite these safeguards, the repossessed goods generally sell for less than their actual value at auction.[36] The buyer in a secured transaction, however, has a significant means of avoiding a sale by the creditor—the buyer can raise cash by selling the property privately. A lessee does not have this option

---

[31] NAT'L CONSUMER L. CENTER, UNFAIR DECEPTIVE ACTS AND PRACTICES, *supra* note 22; R. ABRAMS, *supra* note 25, at 8.

[32] U.C.C. §§ 9-501 to 9-505 (1987).

[33] U.C.C. § 9-503(3) (1987). *See* Ingersoll-Rand Fin. Corp. v. Miller Min. Co., 817 F.2d 1424, 1427 (9th Cir. 1987); Mack Fin. Corp. v. Crossley, 209 Conn. 163, 164, 550 A.2d 303, 304 (1988); BJL Leasing Corp. v. Whittington, Singer, Davis, 204 N.J. Super. 314, 321, 498 A.2d 1262, 1265 (N.J. Super. Ct. App. Div. 1985).

[34] U.C.C. § 9-503(3) (1987).

[35] "[W]hen sufficient notice of sale is not given a presumption arises that the collateral is worth at least the amount of the debt." Franklin State Bank v. Parker, 136 N.J. Super. 476, 482, 346 A.2d 632, 635 (Union County Ct. 1975).

[36] NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES, *supra* note 22, § 5.4.4.2.1; R. ABRAMS, *supra* note 25, at 9–10.

because the lessee does not own the goods. Thus, the consumer in a lease transaction neither has the protections afforded a buyer in a secured transaction nor the right to dispose of the goods to mitigate damages upon default.[37]

Some lessors have modified the classic default formula to credit the lessee with a portion of the amount of remaining future lease payments.[38] One common approach bases its default formula on an analogy between leases and installment loans.[39] This loan model allocates portions of each monthly lease payment to the cost of depreciation of the leased property, the lease charge, and the sales tax.[40] The allocations to the cost of depreciation are roughly analogous to amortization of the principal of a loan; the allocations to lease charges are roughly analogous to interest. While the total amount of each payment is constant throughout the lease, the portion of each payment allocable to the cost of depreciation and lease charge differs with each payment just as the portion of a loan payment allocable to principal and interest differs with each payment on an installment loan. The lease charge, like an interest charge, is a constant percentage of the unamortized balance of the cost of depreciation. As the unamortized balance declines throughout the lease term, the amount of the lease charge declines, and the portion of the payment allocated to amortization of the balance increases. If the lessee terminates the lease early, the lessee owes the unamortized balance of the cost of depreciation but not any future lease charges.[41] The result is analogous to the prepayment of an installment loan where the borrower owes the principal but not

---

[37] *But see* W.L. Scott, Inc. v. Madras Aerotech, Inc., 103 Idaho 736, 741, 653 P.2d 791, 796 (1982) (lessor required to mitigate damages upon default by disposing of property in commercially reasonable manner and reduce future payments to present value); *In re* United Am. Fin. Corp., 55 B.R. 117, 119 (E.D. Tenn. 1985) (discounting to present value lessor's recovery of accelerated balance of future rental payments).

[38] The Rule of 78 often is used to calculate unearned interest in automobile leases. *See* NAT'L CONSUMER L. CENTER, TRUTH IN LENDING, *supra* note 22, at 115; R. ABRAMS, *supra* note 25, at 7. For an explanation of the Rule of 78, see *supra* note 23.

[39] An example of the loan-model damage formula is found in the proposed New York Motor Vehicle Retail Leasing Act, S. 3612 and A. 5924, 1989-90 Leg. Sess. New York, § 341 [hereinafter New York bill]. This proposed act was committed to the Rules Committee as of October 23, 1990.

[40] *See, e.g.*, Security Pacific Auto Finance Vehicle Lease Agreement Closed End UVDS 1652-0589CA 11Y, Items 8 and 16 [hereinafter Security Pacific lease].

[41] Some lease contracts specify that the lessee owes the adjusted lease balance. NAT'L CONSUMER L. CENTER, TRUTH IN LENDING, *supra* note 22, at 116; Nissan lease, cl. I, *supra* note 22.

any future interest. Leases using this default formula usually require the lessee to pay an additional termination fee.[42]

Despite its improvement over the classic default formula, this loan model has substantial problems. First, the lessor establishes the cost of depreciation to be amortized and the rate of the lease charge. A high cost and a low rate of charge will produce the same level of monthly payment as a lower cost and higher rate of charge. However, since the whole unamortized cost is due on early termination, a higher cost will result in higher early termination damages. Second, leases often provide for an inaccurate calculation of unearned lease charges by use of the Rule of 78 and include early termination charges which may be unrelated to the lessor's actual damages.[43] Third, on early termination, the lessor regains the residual value of the leased property earlier than if the lease continued. Normally, the lessee does not receive credit for the early return of the leased property.[44] Because the loan-model damage formula does not regulate these three areas, it produces early termination charges that are unfair to the lessee.[45]

1. Federal and State Statutory Protections: "Notice and Disclosure" Provisions

Federal and state laws do not limit the excessive profits the lessor receives upon the lessee's default or early termination of the lease contract other than to ban unreasonable default and early termination penalties. These laws reflect the "notice and disclosure" approach to consumer protection borrowed from statutes regulating the sale of goods.[46] This approach assumes

[42] Security Pacific Lease, *supra* note 40, at Item 16; Maryland Lease, *supra* note 22, at Item 35; New York bill, *supra* note 39, §§ 331(6), 341.

[43] *See, e.g.*, Citicorp Lease, *supra* note 23, at Item 17. The New York bill does not regulate the calculation of unearned lease charges, but would require early termination charges to decline through the term of the lease. New York bill, *supra* note 39, § 341.

[44] NAT'L CONSUMER L. CENTER, TRUTH IN LENDING app. O, *supra* note 22, at 114.

[45] The New York bill does not regulate the establishment of the cost of depreciation to be amortized and does not require disclosure of the cost or the rate of the lease charge. Nor does the New York bill require credit for early realization of the residual value of the leased property. New York Bill, *supra* note 39, § 341.

[46] The term "notice and disclosure" is borrowed from D. ROTHSCHILD & D. CARROLL, *supra* note 12, at 244 (the Consumer Protection Act is predicated on "disclosure" and "notice"); *see also* Barber, *supra* note 12, at 1227 (knowledge is key to consumer problems); Consumer Leasing Act of 1976, S. REP. No. 590, 94th Cong., 1st Sess. 2, *reprinted in* 1976 U.S. CODE CONG. & ADMIN. NEWS 431, 432 [hereinafter 1976 LEGISLATIVE HISTORY] (legislative history of the Consumer Leasing Act shows Congress's concern with providing disclosure of information to consumers). While consumer credit legislation sets maximum rates for installment credit, the limits are high. These statutes also require disclosure of numerous contract terms.

that a consumer who is informed about transaction costs can make an informed, self-interested choice.[47] The disclosure of information motivates the consumer to shop for the best deal and fosters competition among merchants in the market to establish ethical standards for commercial behavior.[48]

The Consumer Leasing Act is the primary source of consumer lease protection under federal law.[49] It protects consumers against inadequate and misleading cost disclosure, and miscalculation of the residual value of leased goods. The Act requires the lessor to make consumer lease disclosures in a written statement prior to the execution of the lease agreement.[50] The disclosures must be made clearly and conspicuously.[51] Default and early termination penalties must be specified in the lease.[52] "Unreasonable penalties," even if disclosed, violate the statute.[53] However, the statute gives no guidance as to what constitutes an "unreasonable penalty." The standard default provisions of leases, though they are unfairly favorable to lessors, have not been held unreasonable under the federal standard.[54]

Moreover, the Act does not apply to all consumer leases. Rather, it is limited to leases of personal property entered into by a natural person primarily for a personal, family, or household purpose, providing the lease has a duration of at least four months and the total lease obligation does not exceed $25,000.[55] Many consumer leases have a total lease obligation greater than $25,000. Because the size of deficiency claims is related to the

---

[47] The "notice and disclosure" approach to consumer protection assumes that a free-market economy is the best way to distribute resources and risks in society, but recognizes that imbalances of economic power and knowledge exist between consumers and merchants. The "notice and disclosure" approach attempts to restore the equality of consumer and merchant by requiring dissemination of information about the cost of goods. One commentator has noted:

> Today the essential conditions for competition do not exist, and as a result the consumer has been placed in a disadvantageous position. The classical economists assumed that each industry would always have a large number of sellers, with the result that no individual seller would have sufficient economic power to control prices. However, in reality many industries are dominated by a few large sellers, and typically these dominant concerns act jointly like a monopolist . . . . In short, the cornerstones of competition have been pulled out.

Barber, *supra* note 12, at 1222–23.

[48] D. ROTHSCHILD & D. CARROLL, *supra* note 12, at 244.

[49] 15 U.S.C. §§ 1667–1667e (1988).

[50] *Id.* § 1667a.

[51] *Id.*

[52] *Id.* § 1667a (11).

[53] *Id.* § 1667b (b).

[54] There are no cases as of Oct. 8, 1990 construing the federal standard.

[55] 15 U.S.C. § 1667(1) (1988).

total lease obligation, the federal Act thus excludes leases which are capable of producing the largest penalties upon default or early termination. The reasonableness of the default formula should determine the level of consumer protection, not the value of the contract. A consumer spending more than $25,000 to lease goods is not necessarily more capable of understanding the magnitude of liability upon default and early termination, or more likely to have the economic power to negotiate the terms of the lease, than a consumer spending less money.

The only other provision contained in the Consumer Leasing Act requiring more than disclosure of information protects the consumer against payments at the end of the lease term based on changes in the residual value of the leased goods.[56] This provision applies only to leases containing terminal rent adjustment clauses.[57] A TRAC lease requires the lessee to pay the difference in value between the estimated residual value and the realized residual value of the goods. For these leases, the lessor must make a good faith and rational estimate of the residual value the property will have at the end of the lease. If the estimated residual value and the realized residual value differ, the consumer is liable for an amount no greater than three times the average monthly payment.[58] This limitation does not apply to differences between the two values caused by market fluctuations not reasonably foreseeable by the lessor.[59] In that case, the consumer's liability at the end of the lease is unlimited.[60]

The inadequate protections afforded by the Consumer Leasing Act are not remedied in the state provisions. No state has comprehensive legislation resolving the problems posed by consumer lease contracts.[61] The Massachusetts and Washington

---

[56] 15 U.S.C. § 1667b (a) (1988); *see also* 1976 LEGISLATIVE HISTORY, *supra* note 46, at 433–34.

[57] Since terminal rent adjustment clause leases ("TRAC leases") often are found to be leases intended for security subject to article 9 of the U.C.C., this provision of the Consumer Leasing Act is of dubious value for true leases. *See In re* Tulsa Port Warehouse Co., 690 F.2d 809 (10th Cir. 1982); *In re* Tillery, 571 F.2d 1361 (5th Cir. 1978); Columbus Motor Car Co. v. Textile-Tech, Inc., 68 Ohio Misc. 25, 428 N.E.2d 882 (Franklin Cty. Mun. Ct. 1981).

[58] 15 U.S.C. § 1667b (a) (1988).

[59] 1976 LEGISLATIVE HISTORY, *supra* note 46, at 436.

[60] The protection afforded by this provision is not related to the more common and important problem of the measure of damages upon default.

[61] The Uniform Consumer Credit Code ("U.C.C.C."), a uniform state law, regulates consumer credit transactions including consumer leases. U.C.C.C. §§ 1.101–9.103 (1974). Enacted in two versions, the 1968 and the 1974 Acts, it is designed to replace non-uniform state laws dealing with consumer credit. UNIF. CONSUMER CREDIT CODE ACT, 7 U.L.A. 579 (1968); UNIF. CONSUMER CREDIT CODE ACT, 7A U.L.A. 1 (1974).

statutes are analogues of the federal law and are inadequate for the same reasons that the federal law is inadequate. The California and Maryland statutes are specific to motor vehicle leases and cannot serve as general consumer protection statutes.[62] Moreover, the California and Maryland statutes primarily depend upon disclosure to attain their objectives and prohibit practices already forbidden by common law.[63]

Consequently, to extend protection to consumer leases and to resolve the problem of damages upon a lessee's default or early termination of the lease, courts rely upon contract law and consumer protection legislation designed primarily for the sale of goods.[64] Neither approach succeeds in regulating damages

---

U.C.C.C. §§ 1.101–9.103 (1968). The U.C.C.C., which focuses upon sales of goods and small loans, has been adopted by nine states. COLO. REV. STAT. §§ 5-1-101 to 5-9-103 (1974); IDAHO CODE §§ 28-31-101 to 28-49-107 (1980 & Supp 1990); IND. CODE ANN. §§ 24-4.5-1-101 to 24-4.6-1-202 (West 1980); IOWA CODE ANN. §§ 537.1101–537.7103 (West 1987); KAN. STAT. ANN. §§ 16a-1-101 to 16a-9-102 (1988); ME REV. STAT. ANN. tit. 9A, §§ 1-101 to 10-401 (1980 & Supp 1990); OKLA. STAT. ANN. tit. 14A, §§ 1-101 to 9-101 (West 1983); S.C. CODE ANN. §§ 37-1-101 to 37-10-106 (Law Co-op. 1985); UTAH CODE ANN. §§ 70C-1-101 to 70C-9-102 (Supp 1990).

The U.C.C.C. applies to consumer leases for a term exceeding four months providing the amount payable under the lease does not exceed $25,000. UNIF. CONSUMER CREDIT CODE ACT § 1.301(14), 7A U.L.A. 43 (1974); UNIF. CONSUMER CREDIT CODE ACT § 2.106, 7 U.L.A. 638 (1968). The provisions that apply to consumer leases require the disclosure of the elements of consumer lease transactions, contain limitations on agreements and practices applicable to consumer leases, limit the lessee's liability at the end of the lease term, regulate insurance, and provide civil remedies and penalties for violations. *Id.* comment § (14). The provisions of the U.C.C.C. do not address the problems identified in Part I of this Article, though certain provisions, like insurance, are related to those issues. The limitation imposed upon maximum charges in the cost of money or credit does not apply to leases. *Id.*

[62] CAL. CIV. CODE § 2985.7 (West 1974 & Supp. 1990); MD. COM. LAW CODE ANN. § 14-2001 (1990).

[63] Pre-judgment garnishment procedures, forbidden by the California statute, violate fundamental principles of due process. Sniadach v. Family Fin. Corp. of Bay View, 395 U.S. 337 (1969).

[64] Courts apply the contract law principle of "unconscionability" to relieve a consumer from unfair or onerous contracts "imposed by a more powerful party." Shell, *Substituting Ethical Standards for Common Law Rules in Commercial Cases: An Emerging Statutory Trend*, 82 NW. U.L. REV. 1198, 1209 (1988). Some states simply have amended existing consumer statutes to cover leases. *See, e.g.*, CAL. CIV. CODE §§ 1770, 1790 (West 1985 & Supp. 1990), § 2985.7 (West 1974 & Supp. 1990). These sections of the California Code govern deceptive practices, warranties, and vehicle leases, respectively. The original versions of these sections did not apply to leases. Given the dearth of consumer leasing legislation, courts also apply consumer credit and deceptive trade practices legislation to leases that qualify as conventional sales. *E.g.*, Sheffield Commercial Corp. v. Clemente, 792 F.2d 282 (2d Cir. 1986) (finding auto lease constituted a sale; court remanded to determine applicability of New York Motor Vehicle Retail Installment Sales Act); Barco Auto Leasing Corp. v. House, 202 Conn. 106, 520 A.2d 162 (1987) (applying retail installment sales act and unfair trade practices act to lease, thus constituting a sale); Barco Auto Leasing Corp. v. PSI Cosmetics, Inc., 125 Misc. 2d 68, 478 N.Y.S.2d 505 (N.Y. Civ. Ct. 1984) (using analogy approach to find lease was sale, court applied provisions of article 2 of U.C.C. to lease).

upon default or early termination. Contract law is too vague to establish clear guidelines for damage clauses in all cases. Furthermore, remedies for sales transactions are not transferable to lease contracts. Consequently, the decisions in these two areas have produced inconsistent results and have failed to provide a logical and unified approach to consumer leases.[65]

2. Common Law Protections

In the realm of contract law, courts use the "unconscionability" doctrine to determine whether a default or early termination formula in a given lease contract is unreasonable.[66] Under this doctrine, a contract may be adjudged unenforceable, in whole or in part, when a significant inequality exists between the parties, and the contract contains terms unreasonably adverse to the weaker party.[67] Consumer lease contracts generally fit this description. They are frequently contracts of adhesion prepared by the lessor to protect its interests.[68] The lessor usually has greater economic power and knowledge of the market than the

---

[65] Consumer legislation covering sales of goods was not applied to lease contracts in the following cases: Keeling v. Ford Motor Credit Co., 314 Md. 311, 550 A.2d 932 (1988); Ford Motor Credit Co. v. Sims, 12 Kan. App. 2d 363, 743 P.2d 1012 (1987); Sellers v. Frank Griffin AMC Jeep, Inc., 526 So.2d 147 (Fla. Dist. Ct. App. 1988). *Compare Keeling, Ford v. Sims, and Sellers with Sheffield*, 792 F.2d at 282, *Barco v. House*, 520 A.2d at 162, *and Barco v. PSI Cosmetics*, 478 N.Y.S.2d at 505 (finding lease was subject to sales law).

[66] *See, e.g.*, John Deere Leasing Co. v. Blubaugh, 636 F. Supp. 1569 (D. Kan. 1986) (early termination damages clause of lease violated common law doctrine of unconscionability, as well as prohibition against unconscionable contracts contained in article 2-302 of the U.C.C.).

[67] *See* Comment, *Unconscionable Contract Provisions: A History of Unenforceability from Roman Law to the UCC*, 42 TUL. L. REV. 193, 196, 201 (1967); Braucher, *The Unconscionable Contract or Term*, 31 U. PITT. L. REV. 337 (1970); Leff, *Unconscionability and the Crowd—Consumers and the Common Law Tradition*, 31 U. PITT. L. REV. 349 (1970); Murray, *Unconscionability: Unconscionability*, 31 U. PITT. L. REV. 1 (1969); Spanogle, *Analyzing Unconscionability Problems*, 117 U. PA. L. REV. 931 (1969).

[68] For a history and analysis of adhesion contracts, see generally Burgess, *Consumer Adhesion Contracts and Unfair Terms: A Critique of Current Theory and a Suggestion*, 15 ANGLO-AM. L. REV. 255 (1986). *See also* Ehrenzweig, *Adhesion Contracts in the Conflict of Laws*, 53 COLUM. L. REV. 1072 (1953); Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629 (1943). Consumer contracts often are adhesion contracts because the contract is drafted for multiple standardized transactions rather than for a single individual, and the bargaining element is presumably absent from the contract formation process. However, not all standard form contracts are contracts of adhesion. Commercial parties developed standard form contracts to reduce transaction costs. Burgess, *supra*, at 257.

lessee.[69] Lease contracts, therefore, are properly subjected to a test for unconscionability.

Unfortunately, courts apply the unconscionability doctrine to correct only the most egregious and overt forms of disparity in the marketplace.[70] In the context of lease contracts, this may occur because it is often difficult for a court to determine the proper measure of damages in the event of a default on the lease. The extent to which default and early termination clauses are unreasonable may not be apparent within the complex pricing structure of the lease contract.[71] Furthermore, a knowledge of economics and accounting beyond the competence of the court may be required to decipher the damages clause. What constitutes a penalty, as opposed to just compensation for loss, is thus not clear. Courts that equate a deficiency claim in excess of the cash price of the goods with a penalty ignore distinctions between leases and sales.[72]

Moreover, the analytical process that courts employ to test for unconscionability allows judges too much discretion.[73] The

---

[69] Kessler, *supra* note 68, at 632:

> Standard contracts are typically used by enterprises with strong bargaining power. The weaker party, in need of the goods or services, is frequently not in a position to shop around for better terms either because the author of the standard contract has a monopoly . . . or because all competitors use the same clauses.

While Kessler's observations drew from large scale industry practice in the 1940's, they continue aptly to describe the present business setting. Standard form contracts account for nearly all consumer contracts. *See* Slawson, *Standard Form Contracts and Democratic Control of Lawmaking Power*, 84 HARV. L. REV. 529, 529 (1971).

[70] *Cf.* R. UNGER, THE CRITICAL LEGAL STUDIES MOVEMENT 7–8 (1986). Unger's discussion of the function of the doctrine of economic duress in the contracts domain applies with equal validity to the principle of unconscionability because both unconscionability and duress affect the freedom of contract principle. *Id.* at 66–71.

[71] *Cf.* Keeling v. Ford Motor Credit Co., 314 Md. 311, 550 A.2d 932 (1988) (leasing can produce up to four times the profit of a sale at term (citing CONSUMER COUNCIL OF MARYLAND, CONSUMER AUTOMOBILE LEASING STUDY (1986)). Liquidated damages upon default may therefore appear high but not so unreasonable as to render the agreement unconscionable.

[72] *Compare Keeling*, 314 Md. at 311 (lease was not retail installment contract even though the total lease obligation exceeded the value of the goods where the contract included several terms exclusive to leases), *with* Sheffield Commercial Corp. v. Clemente, 792 F.2d 282 (2d Cir. 1986) (lease was retail installment contract because the total lease obligation exceeded the value of the goods). Similar inconsistencies would result if the cash price of the goods were to determine the reasonableness of damages upon default. *Cf.* U.C.C. § 2A-504 Official Comment ("This section does not incorporate two other tests that under sales law determine enforceability of liquidated damages, i.e. difficulties of proof of loss and inconvenience or non-feasability of otherwise obtaining an adequate remedy . . . . Further . . . the last sentence of Section 2-718(1), providing that a term fixing unreasonably large liquidated damages is void as a penalty, was also not incorporated.").

[73] *See* Spanogle, *supra* note 67, at 968 (while the concept of unconscionability is not impossibly vague, "the primary problem with this all-purpose weapon is that . . . neither

list of factors a court may resort to in determining whether a particular contract is unconscionable is long.[74] Courts that distinguish between procedural and substantive unconscionability further complicate the analytical process and burden parties attempting to prove a contract unreasonable.[75] Some courts require a showing of both procedural and substantive unconscionability to invalidate a contract. This burden of proof makes it difficult for the consumer to show that the default or early termination damages clause constitutes an unreasonable penalty.[76] Even when a particular formula is found to be unconscionable, the lessor can alter the formula slightly to circumvent the law. As a device to regulate consumer leases, the concept of "unconscionability" is thus too vague in scope and too infrequently used to offer significant protection.

3. State Consumer Protection Statutes

State consumer protection legislation applicable to lease contracts may be divided roughly into two categories: deceptive trade practice acts and retail installment sales acts.[77] The deceptive trade practice acts are modeled after section 5 of the Federal Trade Commission Act, and are commonly referred to as "state FTC acts."[78] These acts prohibit unfair methods of

---

courts, practicing attorneys, nor contract draftsmen can be certain of its applicability in any particular situation"). Because the concept of unconscionability is vague, judges must exercise a large degree of discretion to apply it to a specific contract.

[74] For a sample list of factors, see John Deere Leasing Co. v. Blubaugh, 636 F. Supp. 1569, 1572 (D. Kan. 1986).

[75] *See id.* at 1573. Procedural unconscionability refers to unfair exploitation of the contract formation process, while substantive unconscionability derives from the terms of the contract.

[76] The burden of proof problems are discussed in Speidel, *Unconscionability, Assent and Consumer Protection*, 31 U. PITT. L. REV. 359 (1970).

[77] For a survey of deceptive trade practice acts, see Weston, *Modern Consumer Protection Laws in the United States and Their Impact Upon Industrial Property*, 22 INDUS. PROP. 28, 41 (1983).

[78] Section 5 declares that "unfair methods of competition and unfair or deceptive acts or practices" are illegal. 15 U.S.C. § 45(a)(1) (1988). *See also* Shell, *supra* note 64, at 1200–01, 1209; Note, *Toward Greater Equality in Business Transactions: A Proposal to Extend the Little FTC Acts to Small Businesses*, 96 HARV. L. REV. 1621, 1622 (1983). For purposes of this Article, no distinction is made between state versions of the Federal Trade Commission Act and consumer fraud statutes. The latter serve essentially the same function as the former by prohibiting unfair, deceptive, and unconscionable commercial practices. Seven states have enacted consumer fraud statutes: Arizona, Arkansas, Delaware, Iowa, Missouri, New Jersey, and South Dakota. ARIZ. REV. STAT. ANN. §§ 44.1521–44.1534 (1967 & Supp. 1989); ARK. STAT. ANN. §§ 4-88-101 to 4-88-112 (1987); DEL. CODE ANN. tit. 6 §§ 2511–2526 (1975); IOWA CODE ANN. § 714.16 (West Supp. 1990); MO. ANN. STAT. §§ 407.010–407.305 (Vernon 1990); N.J. STAT. ANN.

competition and unfair or deceptive conduct in commercial transactions.[79] Some statutes provide a laundry list of unfair and deceptive acts to guide the courts.[80] The state FTC acts authorize the state Attorney General to enjoin unfair practices and impose civil penalties.[81] Most statutes give the consumer a private cause of action to recover damages and attorney's fees.[82]

The protection afforded consumer lessees by some of these state FTC acts is limited because these acts may apply only to the sale of goods.[83] Thus, a court must first wrestle with the question of whether the lease at issue constitutes a sale for purposes of that state's act. Only if the court finds that the lease transaction is a disguised sale is the lessee afforded the protections of a similarly situated buyer. Where state FTC acts do

---

§§ 56:8-1–56:8-48 (West 1989); S.D. CODIFIED LAWS ANN. §§ 37-24-1 to 37-24-35 (1986 and Supp. 1990). *See also* Comment, *Consumer Protection: The Practical Effectiveness of State Deceptive Trade Practices Legislation*, 59 TUL. L. REV. 427, 429 n.16 (1984) [hereinafter *Consumer Protection*].

79 *E.g.*, ALASKA STAT. §§ 45.50.471–45.50.561 (Supp. 1989); CAL. CIV. CODE §§ 1750–1785 (West 1985 & Supp. 1989); COLO. REV. STAT. § 6-1-101 (1974 & Supp. 1989); CONN. GEN. STAT. ANN. § 42-110a–q (West 1987 & Supp. 1989); FLA. STAT. ANN. §§ 501.201–501.213 (West 1988 & Supp. 1989); HAW. REV. STAT. §§ 480-1 to 481F-5 (1988); IDAHO CODE § 48-601 (1977 & Supp. 1989); ILL. ANN. STAT. ch. 121 1/2, para. 261 (Smith-Hurd Supp. 1989); IND. CODE ANN. § 24-5-0.5-1 (West 1980); IOWA CODE § 714.16 (West 1979 & Supp. 1989); MD. COM. LAW CODE ANN. §§ 13-101, 13-301 (1983 & Supp. 1989); MASS. ANN. LAWS ch. 93A, § 1 (Law. Co-op. 1985 & Supp 1989); MICH. STAT. ANN. §§ 19.418(1), 19.418(3) (Callaghan 1981); N.Y. GEN. BUS. LAW § 349 (McKinney 1988 & Supp. 1990); N.C. GEN. STAT. § 75-1-35 (1989); OHIO REV. CODE ANN. §§ 1345.01–1345.02 (Anderson 1979); PA. STAT. ANN. tit. 73, § 201-1 (Purdon 1971 & Supp. 1988); R.I. GEN. LAWS §§ 6-13.1-1 to 6-13.1-19 (1985 & Supp. 1988); TEX. BUS. & COM. CODE ANN. § 17.41 (Vernon 1987); VT. STAT. ANN. tit. 9, § 2451 (1984); VA. CODE ANN. §§ 59.1–59.196 (1987 & Supp. 1989).

80 Weston, *supra* note 77, at 41. The question remains as to whether or not these lists of proscribed commercial practices are meant to be exhaustive. *Compare* Chatham Racquet Club v. Commonwealth, 116 Pa. Commw. 55, 541 A.2d 51 (1988) ("catchall provision" in Section 2 of Pennsylvania's consumer protection law "is designed to cover generally all unfair and deceptive acts or practices in the conduct of trade or commerce"), *with* Commonwealth v. Monumental Properties, Inc., 10 Pa. Commw. 596, 314 A.2d 333 (1973) (because consumer protection law includes a provision for civil penalties, it must be construed narrowly; leasing of real property is not within the law's scope).

81 Comment, *supra* note 78, at 430 ("Typically, a state consumer agency or attorney general's office is empowered to investigate and mediate consumer complaints, obtain assurances of voluntary compliance, and seek injunctions or cease and desist orders.").

82 Treble damages are mandatory for prevailing consumers in nine states: Georgia, Hawaii, Louisiana, Massachusetts, New Hampshire, New Jersey, North Carolina, South Carolina, and Texas. Treble damages are allowed to prevailing consumers in nine other states: Alabama, Alaska, Delaware, Montana, Ohio, Pennsylvania, Tennessee, Utah, and Vermont. *See* Comment, *supra* note 78, at 441.

83 *E.g.*, ALASKA STAT. § 45.50.471 (Supp. 1989); COLO. REV. STAT. § 6-1-101 (1974 & Supp. 1989); DEL. CODE ANN. tit. 6, § 2511 (1975); ILL. ANN. STAT. ch. 121 1/2, para. 261 (Smith-Hurd Supp. 1989). These statutes apply explicitly to sales and may not apply to leases.

apply explicitly to leases, they often fail to provide effective remedies for unreasonable default and early termination penalties. The provisions of these acts admittedly provide relief to the consumer if the default or early termination damages formula is deceptive, unfair, or unconscionable. Most default formulas are disclosed in the contract, however, and are therefore unlikely to rise to the level of a deceptive practice. Furthermore, to the extent that state FTC statutes rely upon the doctrine of unconscionability, discussed above, they do nothing more than mandate judicial application of the common law. Finally, while unfairness is conceptually distinct from unconscionability, they are closely related, and neither provides much protection.

Retail installment sales acts have been enacted in almost every state to govern the sale and financing of consumer goods.[84] These acts are similar, and most limit the rate of interest a seller can charge a buyer for the retail purchase of goods.[85] They also contain disclosure requirements and restrictions on repossession of goods.[86] If the act is held to apply to a particular lease, a violation often bars the lessor from recovering the total amount of the deficiency claimed.[87] Specifically, a lessor's failure to comply with the technical requirements of the act may preclude him from enforcing the default or early termination clause. How-

---

[84] *E.g.*, ALASKA STAT. § 45.10.20 (1989); ALA. CODE § 5-19-1 (1981 & Supp. 1989); ARIZ. REV. STAT. ANN. § 44-6001 (1987); CAL. CIV. CODE § 1801 (West 1985 & Supp. 1989); CONN. GEN. STAT. ANN. § 42-83 (West 1987); DEL. CODE ANN. tit. 6, § 4301 (1975 & Supp. 1988); FLA. STAT. ANN. § 520.30 (West 1988); HAW. REV. STAT. § 476-1 (1988 & Supp. 1989); MASS. ANN. LAWS ch. 255D, § 1 (1980 & Supp. 1989); N.Y. PERS. PROP. LAW § 401 (McKinney 1976 & Supp. 1990); N.D. CENT. CODE § 51-13-01 (1982 & Supp. 1987); OHIO REV. CODE ANN. § 1317.01 (Anderson 1981 & Supp. 1988); PA. STAT. ANN. tit. 69, § 1101 (Purdon Supp. 1989); S.D. CODIFIED LAWS ANN. § 54-3A-1 (1969 & Supp. 1989); TEX. REV. CIV. STAT. ANN. art. 5069-6.01 (Vernon 1987); WASH. REV. CODE ANN. § 63.14.010 (1966 & Supp. 1989).

[85] Johnson, *Regulation of Finance Charges on Consumer Installment Credit*, 66 MICH. L. REV. 81, 88 (1967) ("Most retail installment sales acts provide for rate ceilings that decline as the amount of credit granted increases."). *E.g.*, MASS. ANN. LAWS ch. 255D, §§ 1, 9, 11 (Law. Co-op. 1980 & Supp. 1989); MICH. STAT. ANN. §§ 19.416(101), 19.416(107) (Callaghan 1981 & Supp. 1989); N.J. STAT. ANN. §§ 17:16C-1, 17:16C-41 (West 1984 & Supp. 1989); OHIO REV. CODE ANN. §§ 1317.01, 1317.06 (Anderson 1979 & Supp. 1989).

[86] *E.g.*, CAL. CIV. CODE §§ 1803.1, 1803.3, 1812.2 (West 1985); MASS. ANN. LAWS ch. 255D, paras. 9, 22 (Law. Co-op. 1980); MICH. STAT. ANN. §§ 19.416(103), 19.416(114)(c) (Callaghan 1987); N.J. STAT. ANN. § 17:16C-27 (West 1984) (disclosure provisions only); OHIO REV. CODE ANN. §§ 1317.04, 1317.16 (Anderson 1979).

[87] *E.g.*, MASS. ANN. LAWS ch. 255D, § 29 (Law. Co-op. 1980 & Supp. 1989). Other remedies include rescission of the contract, *e.g.*, CONN. GEN. STAT. ANN. §§ 42-83 to 42-100a (West 1987); *see* Barco Auto Leasing Corp. v. House, 202 Conn. 106, 110, 520 A.2d 162, 166, and penalties for violations of the act. *E.g.*, N.J. STAT. ANN. § 17:16C-56 (West 1984); OHIO REV. CODE ANN. § 1317.99 (Anderson 1979).

ever, only lease transactions that resemble a conventional sale are subject to retail installment sales acts.[88] For a lease to qualify as a retail installment contract, the lease must obligate the consumer to pay a sum substantially equivalent to the value of the goods, and must either vest ownership in the consumer at the end of the lease, or provide the consumer with an option to purchase the goods for a nominal sum. If the consumer's financial obligation under the lease is not substantially equivalent to the value of the goods, the act does not apply. A true lease is therefore not subject to a retail installment sales act. A neat categorization of leases according to this logic is illusory, however: court decisions have failed to articulate a consistent response to the question of whether a lease is eligible for treatment as a retail installment contract.[89]

Even when the lease does qualify as a retail installment sales contract, however, a lease transaction does not fit neatly into the legal categories established to control credit sales. Most retail installment sales acts are designed primarily to restrict finance charges and interest rates. Since leases do not contain explicit interest rates in the rental fees, retail installment sales acts have little or no effect on lease contracts. While it would be possible to impute the rate of interest charged under a lease contract, the computation risks equating leases with sales and assuming that any payment of more than the sales price is interest. Such an approach ignores legitimate differences between a lease and a sale. Moreover, the maximum limitations established under these statutes for interest rates in the sale and financing of goods, while a substantive regulation, are set high enough to accommodate the dynamics of a market economy. Retail installment sales acts therefore do not address default formulas in lease contracts.[90]

---

[88] *E.g.*, N.J. STAT. ANN. § 17:16C-1 (West 1984). The term "retail installment contract" applies to a sale of goods, including "any contract for the bailment or leasing of goods by which the bailee or lessee agrees to pay as compensation a sum substantially equivalent to or in excess of the value of the goods, and by which it is agreed that the bailee or lessee is bound to become, or has the option of becoming the owner of such goods upon full compliance with the terms of such retail installment contract." UNIFORM CONDITIONAL SALES ACT, *cited in* I. MARIASH, A TREATISE ON THE LAW OF SALES 808 app. c (1930).

[89] *Compare Keeling*, 314 Md. 311, 550 A.2d 932 (lease with a value greater than the cash price of the goods is not a retail installment contract), *with Sheffield*, 792 F.2d 282 (lease with a value greater than the cash price of the goods is deemed a retail installment contract).

[90] A few states have consumer protection statutes applicable to motor vehicle leases. *See, e.g.*, CAL. CIV. CODE §§ 2985.7–2990 (West 1974 & Supp. 1989); ILL. ANN. STAT.

4. The Uniform Commercial Code

Article 2A of the U.C.C. is a comprehensive model statute directed at lease transactions. Its provisions, however, do not address specific problems posed by consumer leases. In particular, article 2A does not properly direct the shape and scope of default and early termination damages clauses.[91] Article 2A provides the lessor with three alternate methods of computing damages upon the lessee's default, each of which assumes the lessor has repossessed the property. All three methods allow recovery of accrued and unpaid rent as of the date of default, as well as incidental damages.[92] The first method, applicable when the lessor disposes of the goods by a substantially similar lease, also permits recovery of the present value of the difference between the total rent due under the original lease and the total rent due under the new lease, to the extent it covers the same period.[93] The second method, applicable when the lessor disposes of the goods (1) by sale, (2) by a lease not substantially similar, or (3) in some other fashion, allows recovery of the present value of the difference between the total rent due under the original lease and the market rent at the time and place for tender, computed for the same lease term.[94] The third method, applicable when the lessor does not dispose of the goods or cannot dispose of them for a reasonable price, allows recovery of the present value of the total rent due under the lease contract.[95]

---

ch. 121 1/2, paras. 561–586 (Smith-Hurd Supp. 1989); MD. COM. LAW CODE ANN. §§ 14-2001 to 14-2007 (1990); MASS. ANN. LAWS ch. 255B, §§ 1–25 (Law. Co-op. 1980 & Supp. 1989); N.Y. PERS. PROP. LAW §§ 301–306 (McKinney 1976 & Supp. 1990); *see also* NEV. REV. STAT. ANN. § 100.095 (Michie 1986) (applies to commercial leases only). These acts duplicate the provisions of retail installment sales acts, but most of them apply only to leases that are, in effect, secured transactions. They are not tailored to resolve the problems resulting from default or the early termination of leases. Moreover, they apply only to a narrow category of consumer goods, and thus cannot serve to regulate generally all lease contracts.

[91] *See generally* U.C.C. § 2A. *But see* U.C.C. § 2A (Foreword) (1987) ("Article 2A . . . will apply to consumer's rental of automobiles or do-it-yourself equipment on the one hand, and to leases of such items as commercial aircraft . . . and industrial machinery, on the other. The text recognizes the differences between consumer and business leasing, while resting upon concepts that apply generally to any personal property lease transactions.").

[92] U.C.C. §§ 2A-527 to 2A-530.

[93] *Id.* § 2A-527.

[94] *Id.* § 2A-528.

[95] *Id.* § 2A-529.

The default provisions of article 2A correct some but not all defects in the classic default formula. The general requirement that the lessor reduce to present value the remaining payments due under the lease recognizes the economic benefit to the lessor of early payment, and reduces the lessee's liability for damages. Failure to account for the lessee's liability in present value terms was a major objection to earlier default formulas.[96] Article 2A, however, is by no means a panacea for the extensive potential liability a consumer faces under common default or early termination formulas. A defaulting lessee's liability is mitigated only to the extent of the leased good's market value upon sale or re-lease. This valuation formula is inappropriate because, unlike commercial or industrial stock, a rental market does not exist for most used consumer goods. Indeed, the third default alternative does not require the lessor to dispose of the goods, and thus mitigate damages, at all. The general allowance for incidental damages permits the lessor to charge a broad range of fees for default or early termination.[97] Finally, consistent with its emphasis upon freedom of contract, article 2A allows the parties to alter the default formula by agreement. The lease contract itself can thereby override the slight protections afforded consumers by article 2A.[98]

5. Contract Theory and the Free Market

Current federal and state consumer protection legislation, as well as article 2A, derive from and are largely rooted in classical contract theory.[99] Classical contract theory assumes that the parties to a transaction possess equal bargaining power and

---

[96] NAT'L CONSUMER L. CENTER, TRUTH IN LENDING, *supra* note 22, at 113.

[97] *See* U.C.C. § 2A-530.

[98] Article 2A's unconscionability provision merely incorporates the common law principle of unconscionability, the deficiencies of which are discussed above. *See supra* text accompanying notes 66–76. Article 2A contains other modest consumer protection provisions, including the doctrine of unconscionable inducement: that is, where the contract would not have been entered into but for the unreasonable means employed. Miller, *supra* note 12, at 964. Given the gravity of problems generated by consumer leases, however, the existing consumer protection provisions in article 2A are inadequate.

[99] *See* P. ATIYAH, THE RISE AND FALL OF FREEDOM OF CONTRACT 408 (1979) ("The autonomy of the free choice of private parties to make their own contracts on their own terms was the central feature of classical contract law."). For an examination of the failure of classical contract theory fully to explain the realities of a market economy, see also R. UNGER, *supra* note 70, at 60–68 (the classical model is composed of antagonistic principles and counterprinciples, such as the principle of freedom of contract and the counterprinciple that unfair bargains should not be enforced).

knowledge about relevant market considerations.[100] The parties negotiate at arm's length and enter into contracts to fulfill self-interests. Once a valid contract is formed, the bargain is enforceable even if it subsequently develops that the terms clearly favor one party.

Classical contract theory complements and is essential to the workings of a free market economy. The theoretical assumption, of course, is that society as a whole is enriched by an environment conducive to efficient transactions between informed, self-interested, and unfettered decisionmakers.[101] The state intervenes only rarely in the market since regulation impedes this efficient transaction environment. Neither the classical contract theory nor its free market counterpart adequately explains the realities of the marketplace, however. Classical contract theory ignores the use of adhesion contracts, for example, where the contract's formation process does not occur in the sense of two parties at arm's length dickering over terms. Because the market does not encourage the parties to a consumer transaction to bargain for the agreement, and because consumers are often vulnerable to exploitation in the marketplace, it is impossible to rely on classical contract theory in pure form.

While any governmental regulation is anathema to the classical theories of contract and the free market, legislation has nevertheless been enacted to correct for negative externalities of the market.[102] Consumer protection legislation currently ap-

---

[100] For a concise description of the principal features of the free market economy, see P. ATIYAH, *supra* note 99, at 402–03.

[101] Laissez-faire market theory gave way in the 1930's to a neo-classical model based on John Maynard Keynes' justification of government's active participation in the economy. *See* J.M. KEYNES, GENERAL THEORY OF EMPLOYMENT INTEREST AND MONEY (1936); *see also* J.K. GALBRAITH, ECONOMICS & THE PUBLIC PURPOSE 11 (1973); P. ATIYAH, *supra* note 99, at 693 ("[I]n the modern world . . . it is absurd to think of society as regulated by freedom of contract subject only to limited instances of State 'interference.'").

[102] One of the theoretical presuppositions of market economics is relatively perfect competition in the market, including perfect dissemination of knowledge. Consumer protection legislation arose to correct inherent imperfections in the market caused by, for instance, consumers' inability to wield perfect knowledge or equal bargaining power in standardized commercial transactions. While consumer protection legislation cuts against classical contract theory by limiting the parties' freedom of contract, governmental regulation of the market is in fact rather modest because most consumer protection statutes merely regulate extreme forms of commercial misconduct and provide the consumer with information which theoretically restores the market to its ideal state. The Consumer Credit Protection Act, a disclosure statute, and the Federal Trade Commission Act, a regulatory statute, exemplify this approach. CONSUMER CREDIT PROTECTION ACT, 15 U.S.C. §§ 1601–1677 (1970); FEDERAL TRADE COMMSSION ACT, 15 U.S.C. § 45(a)(1) (1971).

plicable to leases places minimal limits upon the parties' freedom of contract, sustaining the major premise of classical contract theory.[103] Although these statutes recognize gross disparities in knowledge and bargaining power between consumers and merchants, they do not impose significant limitations on the relationship between the parties; in deference to the complementary principles of freedom of contract and the free market, these statutes in effect grant the lessor primary and nearly exclusive control over the setting of default and early termination penalties. Market deformities in lease transactions are corrected only indirectly by the means of "notice and disclosure," maintaining the fiction of equality as between consumer and merchant.

B. *Risk of Loss and Insurance Costs*

An equally important term of any lease contract, from the perspective of the consumer-lessee, is the allocation of the risk that the leased goods may be damaged, lost, or destroyed during the term of the lease. For instance, if the goods are destroyed, is the lessee still obligated to pay rent to the lessor? Conversely, is the lessor required to provide substitute goods to the lessee? If the goods are damaged, does the lessee's obligation to pay rent continue during the repair period when the lessee does not have use of the goods? Should the rights and liabilities of the parties be determined according to which party caused the damage, destruction, or loss of the goods? Or should some other standard determine the rights and liabilities of the parties? Neither the common law nor article 2A provides clear and consistent answers to these questions. Nor, for that matter, do federal and state consumer protection statutes deal with the issue.[104]

At common law, if the leased goods were destroyed without fault of either party, the lease was terminated. The common law is not completely clear, however, on whether the lessee is released from rental obligations for the balance of the lease if

---

[103] *See supra* notes 7–19 and accompanying text.

[104] Federal and state consumer protection statutes applicable to leases merely require the disclosure of insurance provided by the lessor or required to be paid by the lessee. THE CONSUMER LEASING ACT, 15 U.S.C. § 1667a(7) (1988); CAL. CIV. CODE § 2985.8(J) (West 1974 & Supp. 1990); MD. COM. LAW CODE ANN. § 14-2002(a)(2) (Michie 1990); MASS. ANN. LAWS ch. 93, § 91(g) (West Supp. 1990); WASH. REV. CODE ANN. § 63.10.040(f) (West Supp. 1990).

neither of the parties is at fault.[105] Article 2A, by contrast, initially places the risk of loss upon the lessee only in the case of finance leases; otherwise the lessor intially bears the burden. Article 2A fails to specify the rights and liabilities of the parties upon damage, destruction, or loss of the goods resulting from the "fault" of one of the parties.[106] Thus, existing law does not provide a sufficiently detailed scheme of rules to govern the myriad issues that arise in allocating the risk of loss. Moreover, to the extent that the parties are free to vary these rules by contract, existing law allows lessors to shift the risk of loss to the lessee.[107] Shifting the risk of loss to the consumer has the potential to increase the total cost incurred under the lease in ways that the lessee is unlikely to have understood or anticipated.[108]

Assume a consumer rents a garden tool for the weekend under a contract that shifts the risk of loss to the lessee. If the tool is damaged in the course of its ordinary use, the consumer must pay to repair the tool even though the consumer was not at fault and the cost of repair is disproportionate to the rental cost. Although the consumer was at least presumptively aware that the lease contract imposed this potential liability, the result nevertheless appears unfair to the extent that it was not within the practical contemplation of the consumer. A consumer who rents a garden tool for a short period of time and a relatively small sum of money may not expect to pay for an expensive

---

[105] *See* E. GODDARD, OUTLINES OF THE LAW OF BAILMENTS AND CARRIERS §§ 33, 123 (2d ed. 1928); *see also* J. STORY, COMMENTARIES ON THE LAW OF BAILMENTS § 417a (9th ed. 1878); Boss, *Panacea or Nightmare? Leases in Article 2*, 64 B.U.L. REV. 39, 97–98 (1984). In the absence of fault, a lessee was liable for the value of use before destruction; the lessor bore the risk of loss in general. If the lessee was at fault, he was liable for damage due to that fault, including damage to the reversion.

[106] U.C.C. § 2A-219 Official Comment ("This section does not deal with the responsibility for loss caused by the wrongful act of either the lessor or the lessee."). *See also* Boss, *supra* note 105, at 97 ("The Code's rules focus on the transfer of the goods to the buyer; once that transfer occurs, absent default by the seller, the code signs off."). The risk of loss rule in article 2A, like most other rules in the article, applies only in the absence of a lease provision on that subject. *See* U.C.C. § 2A-301. "Because consumers lack the knowledge to evaluate the cost of the risk, a rational seller will draft contract terms that shift risks to the consumer." Meyerson, *The Efficient Consumer Form Contract: Law and Economics Meets the Real World*, 24 GA. L. REV. 583, 605 (1990).

[107] *See* U.C.C. § 2A-301. Subject to the terms of §§ 2A-108 and 2A-219, the parties are free to allocate risk as between themselves in the lease contract.

[108] Consumers primarily focus upon the rental cost of the lease. *See* Meyerson, *supra* note 106, at 595 (Consumers "generally understand the central terms, such as price, but do not know of or do not understand many subordinate terms."). As a result, they are unlikely to take account of their potential liability under the lease if the goods are destroyed. This amount, especially in short-term leases, may far exceed the total cost of rental.

repair. The same result obtains in long-term leases such as automobile leases.[109]

On the other hand, allocating the risk of loss to the consumer-lessee when the lessee is at fault appears more equitable. Assume the same set of facts set forth above except that, because the consumer leaves the tool on the lawn overnight, the tool is stolen. Imposing the burden of loss upon the consumer appears less unfair in this example, since the loss was attributable to the consumer's negligence. The two examples suggest that the risk of loss should be allocated according to fault. Allocating risk of loss according to fault in this context, however, is unworkable because article 2A fails to establish a standard of wrongful conduct for the parties.[110] Moreover, distinctions in degree of fault generally are not made in insured losses, and in the context of motor vehicles, states have tended to abandon fault to allocate the cost of damage.[111]

In general, to the extent article 2A assigns the risk of loss to the lessor, the lessee is insulated from liability for damage, destruction, or loss of the goods.[112] Given that article 2A permits allocating the risk of loss between the parties by contract, it is in the lessor's interest, and usually within the scope of the lessor's bargaining power, to shift the risk of loss to the lessee. In addition, the exception under article 2A for finance leases means that in a substantial number of cases the risk of loss is initially allocated to the lessee.[113] Article 2A is of little help in ensuring that the lessee does not bear a disproportionate share of the risk of loss for damaged, destroyed, or lost goods because it allows the parties to decide the issue themselves; in effect article 2A delegates the decisions about allocation of risk and

---

[109] *See infra* notes 114–115 and accompanying text.

[110] *See* U.C.C. § 2A-219 Official Comment, *supra* note 106. Whether the phrase "wrongful act" used in that comment refers to negligence, gross negligence, intentional conduct, or some other standard is unclear.

[111] The difficulty lies in equating fault with negligence. The standard homeowner insurance policy pays many types of claims even if the damage was caused partly by the negligence of the insured. Similarly, determining which party is negligent may be irrelevant in small automobile accidents. These developments show an increasing tendency to recognize accidents as inevitable in the ordinary use of goods.

[112] *See* U.C.C. § 2A-219. *But see* Official Comment to that section, *supra* note 106.

[113] *See* U.C.C. §§ 2A-219, 2A-103(9) Official Comment. In many leases, especially automobile leases, the nominal lessor is a financial entity entirely separate from either the manufacturer or dealer. *See* Closed-End Lease; Maryland Lease; Lease and Go, Inc. Lease, *supra* note 22 (dealer automatically assigns the lease to a bank).

insurance costs to the lessor by failing to correct for the lessor's greater bargaining power.

The availability and extent of insurance coverage complicates the risk of loss issue. Except for those automobile leases which must comply with motor vehicle insurance law, lease contracts generally do not require the lessee to purchase insurance to protect against damage, destruction, or loss of property. The consumer without insurance bears the full economic loss. A consumer who is aware that a lease contract has allocated the risk of loss to the lessee may believe that insurance purchased for the goods provides full coverage against damage, destruction, or loss of the leased property. However, under most leases, the consumer is liable for damages in excess of the amount paid by insurance when the leased property is damaged, destroyed, or lost because the total lease obligation usually exceeds the insured value of the property at any point during the lease contract. For example, in an automobile lease, if the car is damaged and the insurance proceeds are used to repair it, the consumer may still be liable to the lessor at the end of the term for a reduction in the value of the car.[114] Furthermore, many leases treat destruction of the property as equivalent to default. If the car is destroyed, the consumer may be liable for damages in addition to the insurance recovery.[115] In such situations, the consumer is unlikely to have foreseen the financial exposure. Neither article 2A nor any federal or state statute deals adequately with the problems posed by consumer expectations about risk of loss allocation and insurance coverage.[116]

---

[114] Many leases make the consumer liable for any reduction in value due to change in the property not resulting from normal wear and tear. Intuitively, an automobile damaged in a collision and repaired is worth less than one not so damaged.

[115] The damages due on default can often exceed the market value of an automobile. *See* NAT'L CONSUMER L. CENTER, UNFAIR AND DECEPTIVE ACTS AND PRACTICES § 5.4.4, *supra* note 22, at 189. The amount of the consumer's additional liability depends upon the measure of damages used and when the default occurs. *Cf.* Keeling v. Ford Motor Credit Co., 314 Md. 111, 550 A.2d 932 (what distinguishes a lease from a disguised secured finance agreement is that, even where total lease payments will have exceeded the cost of the vehicle to the lessor, the parties have stipulated in advance to a "substantial" residual value of the vehicle at the end of the lease). Thus, an insurance settlement based on the market value of the automobile which has been totally destroyed often is insufficient to meet the obligation to the lessor. In some cases, the additional amount owed can be substantial.

[116] *Cf.* U.C.C. § 2A-218 (allowing the parties to allocate as between themselves the duty to obtain and pay for insurance). The Uniform Consumer Credit Code restricts charges imposed by the lessor for insurance. U.C.C.C. §§ 4.101–4.112 (1974). *Cf.* PA. STAT. ANN. tit. 73, § 2151 (Purdon Supp. 1989) (requiring coverage for a leased automobile to extend to an intoxicated lessee).

C. *Assignment of Leases*

Most lease contracts prohibit the lessee from assigning the lease contract to another person.[117] The effect of this prohibition is to prevent the lessee from avoiding default or early termination damages by assigning the lease to a third party able to meet its terms. Assignment is forbidden even in response to unforeseeable changes in circumstance. If, for example, a lease contract for an automobile prohibits assignment, and prior to the expiration of the contract, the lessee becomes too infirm to drive the car, the only way the lessee can get out of the lease is to terminate early and pay damages for default.

In the case of a secured finance sale, by contrast, the buyer has the option to avoid default or mitigate damages upon early termination by selling the purchased goods and paying off the balance of the debt. A lessee does not have this option because the lessee does not own the leased property. The lessee's right to assign the lease is thus the best equivalent of the buyer's right to sell the goods.

Article 2A permits the lessor to prohibit the transfer of the lessee's interest in the leased property provided the prohibition is written, specific, and conspicuous.[118] If the contract does not contain a prohibition against assignment, article 2A prohibits either party from transferring its interest, under certain circumstances, when the transfer would materially increase the burden or risk on the other party or change the duties of the lessor under the contract.[119] Other statutory law does not even address the issue. In the absence of any law to the contrary, the freedom of contract principle makes it a virtual certainty that the lessor will prohibit assignment of the lease.[120]

---

[117] *E.g.*, Manufacturers Hanover Wheelease, Inc., Consumer Motor Vehicle Lease Agreement Closed End With Purchase Option 35392F-6/89 [hereinafter Manufacturers Hanover lease]; Closed End Motor Vehicle Lease and Disclosure Agreement 1756-6, Item 25 (1989); Lease and Go, Inc. Motor Vehicle Lease Agreement and Disclosure Statement (Closed End) 0447, Item 30 (1988); Lease It Closed End Motor Vehicle Lease and Disclosure Statement LE-8, Item 27 (1989).

[118] U.C.C. § 2A-303(7).

[119] *Id.* § 2A-303(1)(b).

[120] There is no obvious economic benefit to the lessor in allowing the lessee to assign the lease. As a result, standard form leases do not permit lessee assignment. *See, e.g.*, Nissan Motor Acceptance Corp. Lease, *supra* note 22 (provision in lease allows assignment by the lessor, but makes the following statement for the lessee: "I understand that I have no right to assign any of my rights under this lease"). In a separate provision, the lessee agrees to retain possession of the vehicle and not to "rent it out, sell it, or give it away." *Id.*

II. Model Consumer-Lessee Protection Act

The Consumer-Lessee Protection Act is a model statute drafted in response to the inadequacy of existing law to regulate abuses in consumer leases. The Act applies to all consumer leases and seeks to protect the consumer lessee by regulating the substantive content of the lease contract.[121] Further, the Act disallows freedom of contract for certain aspects of the lease transaction,[122] in an effort to put the consumer and lessor in a situation of greater equality within the marketplace. In response to the issues identified in Part I of this Article, the following provisions of the Act directly resolve the problems of existing consumer protection law: (1) a mandatory default formula applies upon a lessee's default or early termination of the lease;[123] (2) in the event of damage, destruction, or loss of the goods during the lease term, the lessor's recovery is limited to the proceeds of insurance;[124] and (3) the consumer is given the right to assign the lease to a third party.[125]

A. *Default and Early Termination Penalties*

The Model Act does not distinguish between a default and any voluntary early termination by the lessee. The same formula governs both eventualities and allows the lessor to recover: (1) "[a]ny payments due at the time of default plus interest at prevailing rates on those payments"; (2) "[t]he present value at the time of default of any payments due in the future"; (3) "[t]he reasonable cost of repossession"; (4) "[a]ny damage allowed by the Act for loss or injury to the leased property"; and (5) the residual value of the leased property reduced to present value at the time of default.[126] The lessor's damages then are reduced by the value of the leased property at the time of default.[127] While the elements of the default formula are similar to those of article 2A, the method used to calculate the value of the

[121] *See* discussion of the definition of "consumer lease," *infra* text accompanying notes 152–172.
[122] Model Act § 4.
[123] *Id.* § 16.
[124] *Id.* § 12 (Alternative B).
[125] *Id.* § 8.
[126] *Id.* § 15(a).
[127] *Id.* § 15(b).

property at default radically departs from the article 2A method and other methods used by the leasing industry.[128]

The Model Act assumes that goods have a higher value at default than they would have at the end of the lease contract. The credit that the Model Act awards the consumer reflects the mathematical value of the leased goods at the time of default. The mathematical value formula does not rely upon the vagaries of resale or re-rental,[129] but rather establishes a uniform value for similar goods insulated from unpredictable factors such as those variables of the marketplace that have made the disposition of goods by sale produce less than the estimated residual value of the property. Because the mathematical value formula also avoids the problem of calculating a market rent where a lease market does not exist, it provides a better basis for calculating the actual value of the property at the time of default than does a sale or rental of the property.

Even if the calculation of value was based upon the sale of the goods and the protections of article 9 were extended to the disposition of goods, the mathematical value method still has an advantage over the auction model.[130] The article 9 method fails to establish a fair disposition price and mainly creates technicalities which do not assure the commercial reasonableness of the sale.[131] Consumers often depend upon the secured party's failure to send the required notices to mitigate their damages or bar the deficiency claim. Similarly, the secured party relies upon compliance with the notice requirements to establish the commercial reasonableness of the sale of goods.[132] Since a low sales price alone is not proof of commercial unreasonableness, the consumer cannot effectively contest the sale of goods

---

[128] *See* U.C.C. §§ 2A-527(2), 2A-528 (1987); R. ABRAMS, *supra* note 25, at 15–18.

[129] Model Act § 16.

[130] Article 9 requires the creditor to notify the debtor of the time and place of a public sale or the date after which private sale may occur. U.C.C. § 9-504(3) (1987). It also requires the creditor to sell the goods if the debtor has paid 60% of the amount owed. U.C.C. § 9-505 (1987). The sale, whether public or private, must be held in a commercially reasonable manner. *E.g.*, U.C.C. § 9-504(3) (1987). However, that a different kind of sale or a sale at a different time would have yielded a better price is insufficient to establish that the sale was held in a commercially unreasonable manner. U.C.C. § 9-507(2) (1987).

[131] While the notice provisions relating to the sale may be of limited practical benefit to the debtor, violation of their terms by the creditor results in a significant monetary penalty. *See* U.C.C. § 9-507(1) (1987).

[132] While notice and reasonableness requirements are conceptually distinct, there are cases in which the same disposition would be reasonable if notice is given but unreasonable if notice is not given. *See, e.g.*, New Jersey Bank v. Green, 145 N.J. Super. 560, 368 A.2d 431 (Dist. Ct. Camden Co. 1976).

when the secured party has complied with article 9.[133] If the protections of article 9 were extended to article 2A, the same problems would result. The mathematical value method, however, eliminates technicalities that secured transactions law has created for determining the value of goods.

The mathematical value of leased goods at the time of default is calculated by a formula set forth in the Model Act. If the value of the goods is set by a purchase option applicable at the time of default, that value is taken as dispositive. If there is no option to establish a value, the value of the leased goods is determined by interpolating between two established values. In a lease without options exercisable during the term of the lease, the established prior value is the cash price of the goods and the established subsequent value is the residual value.[134] The Model Act requires the lessor to set forth the cash price and residual value in the lease contract.[135] The residual value is the option price for the purchase of the property at the end of the lease term; if the lease does not provide this option, the residual value is the retail price of similar property.[136] Locating the hypothetical value of the goods at default becomes a matter of interpolation between the established prior value and established subsequent value. Two options are provided: the straight-line method, and the 125% declining balance method. The latter method accounts for some of the rapid depreciation of the goods in the beginning of the lease term, and thus compensates the lessor for losses resulting from early termination of the lease.[137] The straight-line method applies to the lease contract unless the lessor specifies otherwise.

---

[133] It is very frequently stated that a sale is not made commercially unreasonable by the fact that it yields less than expected or less than another sale might have yielded. *See, e.g.*, United States v. Excellair, Inc., 637 F. Supp. 1377 (D. Colo. 1986); Leasing Service Corp. v. Graham, 646 F. Supp. 1410 (S.D.N.Y. 1986). The common methods of disposition of repossessed goods—auction or sale on bids to dealers—are unlikely ever to be found unreasonable.

[134] The Model Act contemplates interpolation between the closest established prior and subsequent values. If options establish values, those values are used. However, nothing in the act requires options. The only values which must be set forth in the lease are the cash price and the residual value. Consequently, in many cases the cash price and residual value serve as the prior and subsequent values respectively. Model Act § 16.

[135] *Id.* § 16(a).

[136] *Id.* § 16(b).

[137] A 125% declining balance method allows a decline in value for the first month which is 25% higher than that allowed by the straight-line method. However, each month's decline in value will be less than the last. The result is a method which closer approximates the way goods actually depreciate: quickly at first, more slowly afterward.

To illustrate the application of the two valuation methods, assume a four-year lease of an automobile having a cash price of $10,000 and a set residual value of $4,000. If the lessee defaults after the fourth month, the value of the car is $9,375 by the straight-line method, and $9,244 by the 125% declining balance method.[138] Since this example involves a default very early in the lease term, a lessor might believe that either method overstates the value of the car. However, a lessor can control the value assigned to the goods by providing a purchase option in the lease. The purchase option price is used to calculate damages upon default. Economic factors should guarantee that the option prices will be reasonably accurate. A low option price is an advantage to the lessor because it increases damages upon default, but an unrealistically low option price invites the lessee to exercise the option and terminate the lease.[139]

In addition to establishing a mathematical value for the goods, the Model Act creates a uniform mandatory formula for the

---

[138] Using the straight-line method, the car declines in value evenly at $125 per month for the 48 months; for the fifth month the value is $625 less than the cash price. Using the 125% declining balance method, the reduction in value for the first month is 1.25 times the monthly reduction in value under the straight-line method, or $156. For each subsequent month the reduction in value is $156 divided by the $10,000 cash price and multiplied by the value at the beginning of the month. As a result, the reductions in value using this method begin a little higher than under the straight line method but are reduced with each successive month.

The equations for these calculations are as follows:

Where V_d = value at the time of default, and T_d = time of default in months; V_p = prior established value, and T_p = time in months to which that value is applicable; V_s = subsequent established value, and T_s = time in months to which that value is applicable.

STRAIGHT LINE:

$$V_d = V_p - \frac{(V_p - V_s)(T_d - T_p)}{(T_s - T_p)}$$

Where V_{m1} = value at the end of the first month; V_{mn} = value at the end of the nth month; and V_{mn-1} = value at the beginning of the nth month.

125% DECLINING BALANCE:

$$V_{m1} = V_p - 1.25\,\frac{(V_p - V_s)}{(T_s - T_p)}$$

$$V_{mn} = V_{mn-1}\,\frac{V_{m1}}{V_p}$$

[139] Some lease contracts provide that a lessee may terminate the contract and acquire the property by payment of a specified price. These prices are not simply option prices; they are a compound of early termination damages and option prices. *See, e.g.*, Manufacturers Hanover Lease, *supra* note 117. If such prices were considered option prices, they would work against the lessor's interest as a basis for evaluation of the property upon default. The early termination charges thus should be stated separately from option prices.

computation of damages in all consumer leases.[140] Current state and federal legislation allows a variety of formulas to flourish in the marketplace.[141] The existing statutes also rely upon disclosure of the default formula to protect the consumer. By making one formula mandatory for all lease contracts, the Model Act simplifies the measure of damages upon default and establishes certainty in lease transactions for both the consumer and lessor. As a result, the Act greatly reduces the need for litigation. Moreover, the Act's default formula does not depend upon disclosure for its effectiveness. Whether a lessor discloses the formula in the lease contract does not change the amount of damages due upon default. Applying this formula to the example set forth in Part I, a four-year lease of a $10,000 car with a $4,000 residual value and default after the fourth monthly payment of $250, damages are $2,578 or $2,709 depending on the method of valuation used.[142]

B. *Risk of Loss and Insurance Costs*

The Model Act provides two alternatives to risk of loss and insurance.[143] In the first alternative, the lessor bears the risk of loss for all damage, destruction, or loss to the property in the course of its ordinary use,[144] because the lessor is in a better position than the lessee to purchase insurance and protect its

---

[140] Model Act § 15.

[141] Article 2A allows the parties to establish the formula for damages and even to provide for reasonable liquidated damages. U.C.C. §§ 2A-528(1), 2A-504(1) (1987). The only restriction found in federal law is a bar to unreasonable penalties. 15 U.S.C. § 1667b(b) (1988).

[142] The present value of the remaining 44 monthly payments is $9,177.17. This calculation assumes the use of the ordinary formula for present value: present value of a stream of payments is equal to the payment times $[1 - (1 + i)^{-n}]$ divided by i, where i is the interest rate per payment period and n is the number of payments. CHEMICAL RUBBER PUBLISHING CO., STANDARD MATHEMATICAL TABLES 393 (11th ed. 1957). The present value of a single payment is equal to the payment divided by $(1 + i)^n$, where i is the interest rate and n is the number of time periods applicable to that interest rate. *Id.* The present value of 44 monthly payments of $1 at an interest rate of 5/6% per month (10% per year) is $36.71; the present value of 44 monthly payments of $250 is $9,177. The present value of the $4,000 residual value of the car is $2,776.38. The total of the present values of the payments and the residual value is $11,953, which when reduced by the current value of the car, either $9,375 (straight-line) or $9,244 (125% declining balance), is $2,578 or $2,709.

[143] Both options were considered by the Commission, which recommended adoption of the second option, Alternative B. N.J. LAW REVISION COMM'N, *supra* note 5. The first option, Alternative A, is simpler in concept and solves a broader range of problems, but it requires a different method of writing automobile insurance.

[144] Model Act § 12 (Alternative A).

interest in the property. Under this option, the lessee bears risk of loss for damage, destruction, or loss to the property unrelated to the normal uses of the property, such as risk resulting from intentional destruction or gross misuse. When the property is lost or destroyed within the course of its ordinary use, the lease contract is terminated and the lessee has no further obligation under the lease. When the property is damaged in the course of its ordinary use, the lessor must repair the property or provide substitute goods. Because there is no obligation to pay rent during the repair period, the lessor is encouraged to make a prompt repair of the property. While the cost of leases will increase to reflect the price of the lessor's insurance, the benefit to the lessee outweighs the cost increase. This scheme distributes the risk, insures all losses, and reduces the cost of insurance because the lessor is likely to obtain insurance at a lower cost than the lessee.

In the second alternative, the parties are free to allocate risk of loss between them.[145] However, when the consumer lease provides for insurance against risk of loss—as it usually does in long-term automobile leases—special rules apply to limit the liability of the lessee.[146] The lessee's liability is limited to payment of the deductible amount of the insurance policy. The lessor's recovery is limited to the insurance proceeds plus the deductible. When the insured property is damaged and can be repaired, the lessor must select one of two options: (1) apply the proceeds of the insurance policy and the deductible to repair the leased property, or (2) with the consent of the lessee, retain the insurance proceeds and terminate the contract. Thus, the lessor cannot decide to repair the leased property and then, at the end of the contract, demand that the lessee pay an additional sum for reduced residual value. If the lease is terminated, the lessee has no further obligation to the lessor. Damage, destruction, or loss of leased property does not constitute a default entitling the lessor to recover future rental payments, lost profits, or other charges.

Alternatively, if the lessor decides to have the leased property repaired, the lease contract is suspended for the duration of the repair period. The lessee need not make any payments while the lease is suspended. Rather, the lease continues when the

---

[145] *Id.* § 12 (Alternative B).
[146] *Id.*

leased property is returned to the lessee, and the term of the lease is extended for a period of time equal to the duration of the suspension. When the insured property is damaged but cannot be restored to its prior condition, the lessor has no options; it must accept the insurance payment and terminate the contract. The lessee in this situation is obligated to pay only the deductible amount of the insurance policy.

C. *Assignment of Leases*

The Model Act gives the lessee the right to assign the lease to a third party,[147] thus enabling the lessee to avoid an anticipated default. A lessee who cannot afford the rental payments or whose changed circumstances make the lease burdensome would also be able to avoid default by assigning the lease contract. The lessor can disapprove an assignment, but must set forth specific facts that show the assignment actually increases the risk to the leased property.[148] An assignment of the lease is not a factor that alone increases risk to the leased property. Rather, the lessor must articulate facts—other than the assignment itself—demonstrating that its disapproval of the assignment is reasonable. This requirement is intended to facilitate the assignment of lease contracts by the lessee while at the same time protecting the interests of the lessor. If the lessor's disapproval of the assignment is unreasonable, the lessee has the option to terminate the lease contract without further obligation to the lessor.

The Model Act includes other provisions as well.[149] For example, the Act corrects the failure of the Uniform Commercial Code article 2A to protect the consumer against the disclaimer of warranties. Article 2A permits the lessor to exclude both express and implied warranties so long as the disclaimers are conspicuously set forth in the contract.[150] Article 2A sanctions

---

[147] *Id.* § 8.
[148] *Id.* § 8.
[149] Other provisions in the Model Act reproduce with some modifications consumer protection provisions found in existing state law and apply them to consumer leases. For example, the provision making the original lessor and subsequent assignee liable for all claims and defenses arising under the contract is borrowed from the U.C.C.C. U.C.C.C. § 3.404 (1974). Unlike the U.C.C.C., however, the consumer can assert the claim or defense directly against the assignee, without first having to obtain satisfaction from the original lessor.
[150] U.C.C. § 2A-214 (1987).

the disclaimer of oral promises made to induce a consumer to enter into a lease transaction. The Model Act, however, prohibits the disclaimer of any express or implied warranty made to the consumer by making these warranties irrevocable.[151]

III. Conceptual Differences Between the Model Act and Existing Consumer Protection Legislation

Not only does the Model Act resolve the most serious practical problems raised by consumer lease contracts, but its provisions also highlight some of the conceptual weaknesses of current federal and state consumer legislation. One conceptual advantage of the Model Act is its approach to the definition of the term "consumer lease." A consumer lease is generally defined as a lease of goods by a natural person primarily for a "personal, family or household purpose" from a person in the business of selling and leasing goods.[152] However, though the phrase "personal, family or household purpose" is the common denominator of most statutory definitions of the term "consumer lease," several variations in other parts of the definition cause the statutes to differ. As a result, the statutes do not recognize a uniform definition of the term "consumer lease." In addition, no clearly identifiable class of beneficiaries entitled to consumer law protections is apparent within the present statutory framework. Numerous practical problems also exist because of the variations among the statutes. More seriously, the theoretical concept of "personal, family or household purpose" is fundamentally flawed.

The main differences in the definitions of the term "consumer lease" pertain to the duration of the lease term, the maximum value of the lease contract, and the option to purchase the goods at the end of the term. For example, the federal Consumer Leasing Act applies only to a consumer lease with a minimum

---

[151] Model Act § 6.

[152] The phrase "personal, family or household purpose," which is used to distinguish consumer from commercial transactions, derives from the article 9 definition of "consumer goods." U.C.C. § 9-109 (1987). The final draft of article 9 published in 1950 contains one of the earliest references to this phrase and defines consumer goods as those "used for the debtor's personal, family or household purposes." U.C.C. § 9-109(1) (Proposed Final Draft 1950). Almost every consumer statute regulating the sale of goods or credit uses the phrase "personal, family or household purpose" to define a consumer transaction. *E.g.*, U.C.C. § 2A-103(e) (1987); 15 U.S.C. § 1667 (1988); Magnuson-Moss Warranty Act, 15 U.S.C. § 2301 (1988).

duration of four months and a maximum total obligation of $25,000.[153] On the other hand, the Maryland Motor Vehicle Leasing Act applies only to leases with a duration of more than 180 consecutive days and does not contain any maximum financial limitation.[154] Moreover, the official text of article 2A contains a $25,000 limitation similar to federal law, but does not have a four-month minimum requirement for the duration of the lease.[155] Thus, a single lease transaction in Maryland for a motor vehicle lease with a six-month duration and a maximum value of $15,000 would be covered by the Maryland statute, article 2A, and the Consumer Leasing Act, while a similar motor vehicle lease with a value of $30,000 would be covered only by the Maryland statute.[156] Countless other similar examples pervade the law.[157]

These distinctions between the statutes lack a valid basis. As the prices of goods increase, financial limitations on the value of consumer lease contracts become obsolete and constantly need revision if they are to remain meaningful. These "maximum obligation" limits are also arbitrary, since some items that ordinarily are used for a personal, family, or household purpose may exceed the financial threshold. The establishment of "minimum duration" requirements for lease contracts to qualify for coverage under the statutes is also unsound. A consumer who rents a garden tool for the weekend deserves—from a theoretical standpoint—the same protection as a long-term lessee, even though the gardener may have a smaller economic stake in the lease. Consumers with short-term leases are entitled to disclosure of terms because most leases, whether short-term or long-term, shift the risk of loss to the lessee and increase the consumer's potential liability.

State FTC acts that recognize corporations as consumers further complicate the concept of what constitutes a consumer

---

[153] 15 U.S.C. § 1667(1) (1988).

[154] MD. COM. LAW CODE ANN. § 14-2001(c)(i) (1990).

[155] U.C.C. § 2A-103(1)(e) (1987).

[156] This assumes adoption of the official text of article 2A in Maryland.

[157] For example, assume a slightly different hypothetical where the lease term is for four months and the value of the lease exceeds $25,000. In California, the lease is subject to the California Vehicle Leasing Act but not the Consumer Leasing Act. In Maryland, the same transaction is not subject to any statute. The differences in treatment are not justified by legitimate economic or policy reasons, but, rather, are most likely the result of piecemeal legislation.

lease.[158] While it makes sense to extend protections against illegal business practices to corporations, including corporations within the definition of "consumer" obfuscates the concept of consumer for the law generally. Except for these statutes, consumer protection legislation generally does not recognize corporations as consumers. This recognition is inconsistent with the "personal, family or household purpose" test that unifies the statutes in the consumer law area.

The basic concept of a consumer as one who leases primarily for a "personal, family or household purpose" is theoretically inadequate for identifying a comprehensive approach to consumer lease transactions. The genesis of modern consumer protection legislation parallels the development of mass production industry and the use of standard form contracts.[159] Because consumer contracts were not subject to negotiation, consumers were unable to negotiate contract terms with the sellers of goods. This inability of the consumer spurred legislatures to enact consumer protection statutes.[160] The relatively unequal economic power and knowledge of the market between the consumer and merchant were the basis for extending special protections to the consumer.

The small business entity is often no better off economically in relation to the lessor than the individual.[161] Like the individual, the small business owner usually lacks the ability to negotiate the terms of the lease contract.[162] Because the relatively unequal economic power of the individual consumer and merchant—and the use of adhesion contracts—supported the extension of special protections to the individual consumer, similar protections should extend to any party who cannot negotiate

---

[158] The growing use of state FTC acts by businesses in conventional litigation is explored in Shell, *supra* note 64, at 1198 (1988); *see also* Bragg, *Now We're All Consumers! The 1975 Amendments to the Consumer Protection Act*, 28 BAYLOR L. REV. 1 (1976); Comment, *The Deceptive Trade Practices-Consumer Protection Act: The Shield Becomes a Sword*, 17 ST. MARY'S L.J. 879 (1986).

[159] D. ROTHSCHILD & D. CARROLL, *supra* note 12, at 6–8; *see also* P. ATIYAH, *supra* note 99, at 544; Burgess, *supra* note 68, at 259 (linking the spread of adhesion contracts to consumer mass marketing); Kessler, *supra* note 68, at 629.

[160] This follows the general development pattern of consumer legislation. Usury laws were conceived and drafted to meet the problems of loan sharks. *See* Johnson, *supra* note 85, at 82. Laws regulating finance charges were designed to fill the gap left open by usury laws which did not apply to installment sales credit because of the time-price doctrine and so on. *Id.*

[161] The basic outlines of this argument are set forth in an article advocating extension of state FTC acts to small businesses. Note, *supra* note 78, at 1621.

[162] *Id.* at 1629.

the terms of the contract.[163] The policy considerations underlying the traditional consumer protection statutes thus apply equally to small business entities, and the extension of consumer protections to small businesses is the logical result. The definition of "consumer lease," therefore, should focus not on the identity of the lessee or on the anticipated use of the goods, but rather upon whether the lessee had a genuine opportunity to bargain for the agreement.

The arguments against extending consumer protections to business entities are less persuasive.[164] While some small businesses may have bargaining advantages over large corporations because of their expertise in a particular field, the majority of small companies lacks knowledge of the market.[165] Similarly, entrepreneurs may do more product research and comparative analysis of contract terms than the average individual given the commercial character of the transaction. Like the individual, however, small businesses cannot appreciate the consequences of many provisions of the contract without a lawyer. They also may not have the leverage to negotiate those terms of a contract which they think unfair or otherwise do not like.[166] Finally, consumer statutes which award attorney's fees enable small businesses to litigate valid claims that otherwise would be precluded. To reason that large corporations must take the cost of litigation into account and small businesses should do likewise does not justify excluding small businesses from the definition of consumer.

The Model Act expands the definition of the term "consumer" to resolve the practical and theoretical problems of the "personal, family or household" definition. The term "consumer lease" adopted by the Model Act applies to three classes of lease contracts.[167] First, a consumer lease is any lease between

---

[163] *Id.* at 1631.

[164] One commentator has criticized the rationale of the "small business as consumer" analogy to oppose extension of state FTC acts to small businesses. Shell, *supra* note 64, at 1198, 1237. That author contends that open-ended standards of state FTC acts "combined with liberal awards of treble damages and attorneys' fees create incentives for litigation that are inappropriate in the commercial setting." *Id.* at 1253–54.

[165] Note, *supra* note 78, at 1629.

[166] Professor Unger notes "[g]ross inequalities of bargaining power . . . are all too common in the current forms of market economy, a fact shown not only by the dealings between individual consumers and large corporate enterprises, but also by the huge disparities of scale and market influence among enterprises themselves." R. UNGER, *supra* note 70, at 70.

[167] *See* Model Act § 3(a).

a lessor regularly engaged in the business of leasing and selling personal property and a lessee who is a natural person, provided that the property is normally used for a personal, family, or household purpose.[168] The phrase "normally used for personal, family or household purposes" denotes that it is the objective character of the personal property, not the subjective intention of the user, that determines whether a lease is a consumer contract.[169] Second, a consumer lease is any lease between a lessor and a lessee who is a natural person, provided that the total of periodic payments under the lease is $50,000 or less.[170] Third, a consumer lease is a lease between a lessor and lessee that is a business entity, provided that the total of periodic payments due under the lease is $50,000 or less and a natural person is liable for damages and performance of the lease contract.[171]

This third class of consumer lease transactions covers businesses. If an individual must risk his or her personal assets to secure a lease for the business entity, then the lease is more like a consumer lease than a commercial one, and the lessee—the business entity—is entitled to consumer protection.[172]

The other major conceptual difference between the Model Act and most existing consumer law applicable to lease contracts is the reduced reliance upon disclosure to correct the market imperfections related to consumer transactions.[173] The "notice and disclosure" approach appears flawed for four reasons. First, it presumes a model consumer who probably does not exist in the marketplace.[174] Consumers rarely read the dis-

[168] *Id.* § 3(a)(1).

[169] The Magnuson-Moss Warranty Act also uses the word "normally" in conjunction with the phrase "personal, family or household purposes." 15 U.S.C. § 2301(3) (1988). *See also* Donnelly & Donnelly, *Commercial Law*, 37 SYRACUSE L. REV. 307, 332 (1986).

[170] Model Act § 3(a)(2).

[171] *Id.* § 3(a)(3).

[172] The personal guarantee requirement includes letters of credit, surety bonds, or any other device that makes a natural person risk personal assets to obtain the lease in the name of the small business. This definition has several advantages over traditional definitions of a small business, which usually focus on the commercial value of the business or the number of its employees. Given the constant evolution of the market, each of the traditional statutory methods of defining a small business rapidly becomes obsolete and requires periodic legislative amendment. Second, the traditional cutoff points used to distinguish a small business are inevitably arbitrary, no matter how carefully drawn.

[173] The reference is to the Consumer Leasing Act, state FTC acts, state retail installment sales acts and other legislation governing consumer lease transactions.

[174] Consumer ignorance of the market is a well-established fact. P. ATIYAH, *supra* note 99, at 621; Barber, *supra* note 12, at 1227; Johnson, *supra* note 85, at 91. The increasing complexity of products and lease contracts makes it unlikely that more information will reduce consumer ignorance of the market.

closure statements set forth in the lease contract and do not make decisions to lease goods based upon this information. Even if consumers did read the disclosure statements, there is no reason to believe that they would either understand the statements or forego a contract because it contained unfavorable terms. The enactment of disclosure statutes has not stopped consumers from making contracts against their self-interest. Moreover, while disclosure of a single important piece of information—such as the percentage rate charged in a loan—may have some effect upon the consumer, provisions requiring disclosure of subtle contingent matters—such as the measure of damages on default—probably have none. Additionally, each new requirement reduces the prominence of any particular disclosure. Disclosure statutes have resulted in longer and more complicated contracts that the average consumer cannot comprehend.

Second, the supply and demand principle implicit in the "notice and disclosure" approach does not apply effectively to leases. Consumer statutes set maximum finance rates which lower the cost of financing consumer transactions. However, these rate ceilings usually are transformed into the prevailing rate, and variations in the cost of goods purchased on credit are eliminated. To the extent that provisions on maximum finance charges do not apply to consumer leases, the "notice and disclosure" approach does not address the factors underlying excessive costs in consumer lease transactions. If applied to leases, the supply and demand principle would not produce different results.

Third, large corporate enterprises use standardized lease contracts which contain few protections for the individual consumer.[175] These contracts do not vary from company to company as to the kinds of issues raised here.[176] A consumer often cannot negotiate the terms of the lease other than the price and the method of payments, since gross inequalities of power exist between the parties. The comparative shopping for best terms that the "notice and disclosure" approach requires of the consumer cannot take place in a market that relies upon contracts of adhesion to facilitate lease transactions. Competition has

---

[175] Ehrenzweig, *supra* note 68, at 1072; Kessler, *supra* note 68, at 629.

[176] *Compare* Nissan lease, *supra* note 22, *with* General Motors Acceptance Corporation lease, *supra* note 22.

failed to produce significant variation in the terms contained in lease contracts.

Fourth, enforcement of these statutes depends partly upon the consumer discovering lease disclosure violations.[177] Since the statutes do not place any controls upon what practices are allowed in the market, but solely rely upon consumers and governmental agencies to discover disclosure violations in the lease contract, they countenance abuses in the consumer leasing industry until the specific practice is deemed illegal. The "notice and disclosure" approach shifts the onus of enforcement of trade regulation to the consumer because direct regulation would conflict with the free-market economy that the statute is intended to preserve. Whatever merits this approach had in an earlier economic period toward fostering industrial growth are not relevant in the present market.

A recent effort to support the "notice and disclosure" approach uses economic theory to formulate solutions to the problem of consumer form contracts.[178] The approach called the "doctrine of reasonable expectations" focuses upon the reasonable expectations of the consumer, and enforces "those form contract terms that are entered into knowingly and voluntarily by consumers with adequate information."[179] The final contract consists of three types of terms: (1) explicitly agreed-upon terms, (2) terms the consumer is led to believe are contained in the contract, and (3) terms which rational parties would have negotiated.[180] The final contract is supposed to approach the ideal agreement voluntarily entered into by parties with perfect market information. An analysis of this doctrine demonstrates not only the need to abandon permanently the "notice and disclosure" approach to consumer protection, but also reveals the poverty of economic analysis theory to generate a cogent solution to the problem of consumer contracts.

The "doctrine of reasonable expectations" considers contract terms the subject of explicit agreement when they are orally

---

[177] For example, the Consumer Leasing Act gives consumers a private cause of action against lessors for disclosure violations. 15 U.S.C. § 1667d (1988). A consumer can recover statutory damages up to $1,000, actual damages, and attorney's fees. 15 U.S.C. § 1640 (1988). Class actions are limited to $500,000 or one percent of the lessor's net worth. *Id.*

[178] Meyerson, *supra* note 106, at 611. For an explanation of the descriptive and normative aspects of economic analysis of law, see, e.g., R. POSNER, ECONOMIC ANALYSIS OF LAW (2d ed. 1977).

[179] Meyerson, *supra* note 106, at 612.

[180] *Id.*

explained to the consumer rather than merely printed in the contract.[181] To require oral explanations for every significant term of the contract is totally unworkable, and the consumer would not necessarily better understand the contract after the merchant's lecture. Oral explanations of contract terms would decrease efficiency, and thus increase the cost of goods, because transactions would be slow and cumbersome and would require a better-educated work force. Moreover, explaining contract terms orally does not address the question of adverse terms because it does not allow the consumer to dicker for new terms.

The "doctrine of reasonable expectations" would also enforce advertising representations that conflict with written contract terms, and would shift most risk of loss to the merchant.[182] The former concept is covered adequately by existing consumer protection statutes prohibiting deception in market transactions. Because the merchant may allocate the risk of loss to the consumer by explaining the contract clause, this framework for analyzing consumer contracts does not differ from existing law. To the extent that merchants who elect to absorb the risk of loss pass the cost to the consumer in the form of higher prices, it is not clear that this doctrine maximizes social wealth better than alternative market restrictions.[183]

The fiction of treating consumer contracts as consensual agreements is no longer a useful metaphor for resolving the legal conflicts that follow from standard market transactions. The paradigm of a contract as two individuals consenting to terms of an agreement, and thus democratically establishing private law to govern their transaction, does not aptly fit modern consumer transactions.[184] To sustain this fiction by disseminating information and restoring hypothetical equality to the transaction ignores the reality of the marketplace. The better approach is to establish a public—not private—standard of fairness to govern problematic aspects of transactions between merchants and consumers. The publicly-created standard would reflect the public interest and the legitimacy of contracts would rest entirely upon compliance with the public standard. The Model Act is intended to provide such a publicly-created standard.

---

181 Meyerson, *supra* note 106, at 613.
182 *Id.* at 614, 618.
183 *Id.* at 619.
184 The conception of contract law as private lawmaking power is borrowed from Slawson, *supra* note 69, at 530.

The Model Act creates a nonwaivable statutory contract to govern lease transactions between a commercial lessor and consumer.[185] The statute is not variable by private contract. Although disclosure of terms is required,[186] the Act does not depend upon disclosure to correct marketplace inequities. The Model Act directly controls the relationship of parties to a lease transaction to make certain that the market operates upon principles of fundamental fairness. Although the Model Act does not solve the deep flaws of classical contract theory, the problems of leases are resolved legislatively outside the context of contract law. To this extent, the Model Act thus acknowledges the antagonism of principles contained in classical contract law and constructs a solution to lease problems which harmonizes the interests of the parties.

The Model Act's approach to consumer protection allocates resources and risks between the lessor and consumer according to the actual exchange of values between them. Priority is given to the public interest in providing protection to consumers who must trade in the marketplace for goods in order to satisfy the needs of material existence. It therefore rejects the assumption that the market operates most efficiently without controls. Existing law, which mainly provides consumers with information, creates an illusion of equality between lessors and consumers in the marketplace. In an age where market transactions are not vindications of personal freedom, the Model Act introduces ethical standards of conduct and trust into the market and establishes equality between the consumer and lessor.

III. CONCLUSION

The Consumer-Lessee Protection Act establishes a comprehensive framework to govern consumer lease transactions. The policy considerations supporting the Model Act are similar to those supporting the Uniform Commercial Code: simplifying the law and enhancing the certainty of legal standards. The Consumer-Lessee Protection Act thus accomplishes for consumer lease transactions what article 2A accomplished for commercial lease transactions. The Model Act is designed to generate dis-

---

[185] Model Act § 4.

[186] Disclosure of terms is required for leases not governed by the Consumer Leasing Act. 15 U.S.C. § 1667a (1988).

cussion of the practical and theoretical problems in consumer leases and to serve as a starting point for a uniform state act. Economic development of the lease thus has produced a corresponding development in the legal system.

APPENDIX

CONSUMER-LESSEE PROTECTION ACT

SECTION 1. Title.

This Act shall be known and may be cited as the "Consumer-Lessee Protection Act."

SECTION 2. Findings.

The Legislature finds that consumer lease contracts account for a large percentage of consumer transactions. Most consumer lease contracts contain provisions that are unfair to the consumer. Individual consumers generally have less economic power than lessors and cannot negotiate the terms of the lease contract. The terms of the lease contract are established by the lessor and submitted to the consumer on a "take it or leave it" basis. Consumer lease contracts are therefore contracts of adhesion. Existing law does not protect lessees adequately. This legislation establishes standards of conduct in the marketplace for consumer lease transactions.

SECTION 3. Definitions.

a. A "consumer lease" is a lease of personal property between a lessor regularly engaged in the business of leasing or selling and:

1) a lessee who is a natural person and the personal property is normally used for personal, family or household purposes; or

2) a lessee who is a natural person, where the total of periodic payments due under the lease is $50,000 or less, excluding payments for options to renew or buy; or

3) a lessee which is a business entity, where the total of periodic payments due under the lease is $50,000 or less, excluding payments for options to renew or buy, and the lease provides that a

natural person is liable for performance of the obligations of the lease and damages upon default.
b. "Present value" is the amount as of a date certain of one or more sums payable in the future, discounted to the date certain. The discount is determined by a commercially reasonable rate that takes into account the facts and circumstances of each case at the time of the transaction.

SECTION 4. Waiver; agreement to forego rights.

Any term of a lease agreement inconsistent with the provisions of this Act, and any waiver of the protections of this act is unenforceable.

SECTION 5. Three-day grace period; refund of payment.

a. A lessee has the right to cancel an executed lease contract within three business days from the date the lease contract is executed, provided the lessee has not taken possession of the property.
b. Any payment made by a lessee to a lessor pending the execution of a lease contract shall be refunded to the lessee in the event the lease contract is not executed. Any payment made by a lessee to a lessor, whether before or after the execution of a lease contract, shall be refunded if the contract is cancelled pursuant to subsection (a).
c. The lessor shall give written notice to the lessee of his rights under this section.

SECTION 6. Warranties.

In a consumer lease, a disclaimer of any warranty is unenforceable.

SECTION 7. Liens.

Any provision in a consumer lease that gives the lessor a lien on property, other than the leased property or a security deposit, is unenforceable.

SECTION 8. Assignment of consumer leases.

A lessee shall have the right to assign a consumer lease with a term of one year or more provided that the lessor is given written notice of the proposed assignment and does not disapprove it. The lessor shall not disapprove the assignment unless the lessor gives written notice of the disapproval to the lessee within thirty days of notification of the proposed assignment. The notice of disapproval shall contain a statement of reasons showing that the proposed assignment increases the risk to the leased property. After an assignment, the original lessee and the assignee shall be jointly obligated under the lease.

SECTION 9. Assignee subject to claims and defenses.

An assignee of the lessor's rights is subject to all claims and defenses of the lessee against the lessor arising from the lease limited only by the amount of the lessee's total payments under the lease.

SECTION 10. Liability of dealers and remote lessors.

In a finance lease, in addition to the lessor named on the lease, a person who negotiates the lease with a consumer lessee is a lessor for purposes of this Act.

SECTION 11. Specificity of payment terms.

a. For any "consumer lease" defined in this Act that is not subject to the federal regulations regarding disclosure of lease terms, the lessor shall state the date any payment is due and shall:
1) specify the amount of the payment, or
2) provide a formula which allows the amount to be calculated arithmetically.
b. A requirement that makes the lessee responsible for damage to the leased property shall not be construed to be a violation of subsection (a) of this Section, and shall be permissible to the extent allowed by Sections 12 and 13 of this Act.

SECTION 12.(Alternative A) Risk of loss.

The lessor bears the risk of loss of the leased property for any risk resulting from the ordinary use of the property.

SECTION 12.(Alternative B) Risk of loss; Insurance.

a. The lessor bears the risk of loss of the leased property unless the lease specifies in clear and comprehensible language the nature and extent of the risk allocated to the lessee.
b. When the lease allocates a risk of loss to the lessee, provides for insurance against this risk, and the lessee possesses the property, subsections (c) through (f) apply. However, these subsections do not apply if the lease requires the lessee to carry insurance, and the lessee fails to comply with the requirement.
c. If the leased property is damaged, and can be restored to its condition prior to damage, the lessor shall elect one of the following options:
1) apply the amount of the damage as determined by the insurance company (the proceeds of the insurance plus any deductible amount as provided in the insurance policy owed by the lessee) to repair the leased property, and continue the lease, or
2) with the consent of the lessee, retain the amount of the damage as determined by the insurance company and terminate the lease contract.
d. The lease shall be suspended, and the lessee need not make any required payments during the period that the leased property is repaired pursuant to this section. The lease term shall be extended for a period equal to the period of suspension.
e. If the leased property is damaged, and cannot be restored to its condition prior to damage, the lessor shall retain the amount of the damage as determined by the insurance company and terminate the lease.
f. Damage or loss to the leased property does not constitute a default on the part of the lessee, and if the lease is terminated pursuant to this section, the lessor may not recover future rental payments, lost profits, penalties, or other charges.

SECTION 13. Late fees.

a. A late payment fee of no more than five percent of the monthly payment in default, or the sum of $5.00, whichever is less, may

be charged by the lessor for the lessee's failure to make a payment on time.
b. A payment is made on time if made within 10 days of the due date set by the lease contract.
c. Any late payment fee not claimed by notice in writing within 40 days from the date of default is waived.

SECTION 14. Notice of consumer's right to cure.

a. After a lessee has failed to make a required payment for 10 days, the lessor may declare a default by giving the lessee written notice of the default and the right to cure the default. The notice shall contain: the name, address, and telephone number of the lessor to whom payment should be made; the amount of the payment; a statement of the right to cure the default; and the date by which the payment must be received to cure the default.
b. For 20 days after the notice is given, the lessee may cure all defaults consisting of a failure to make a required payment by paying all unpaid sums due at that time.
c. If the lessee does not make payment within the time allowed to cure the default, the lessor may exercise his or her rights under the law.

SECTION 15. Default by lessee.

a. If the lessee defaults or wrongfully terminates a consumer lease, the lessor may cancel the lease, repossess the leased property, and recover no more than the following damages:
1) Any payments due at the time of default plus interest at prevailing rates on those payments;
2) The present value at the time of default of any payments due in the future;
3) The reasonable cost of repossession;
4) Any damage allowed by the Act for loss or injury to the leased property; and
5) The value of the leased property at the end of the lease term reduced to the present value as of the time of default.
b. The lessor's damages are reduced by the value of the leased property at the time of default.

SECTION 16. Calculation of value of leased property.

For determination of damages upon default by a lessee, the value of leased property shall be determined in the following manner:
a. The value at the beginning of the lease term is the retail sales price of the leased property. This price shall be stated in the lease contract.
b. The value at the end of the lease term is the option price established for the purchase of the leased property at the end of the lease term. If no option price is stated in the lease, the value is the average retail market price for similar property.
c. The value at any other time within the lease term is the option price established for the purchase of the leased property exercisable at that time.
d. If the value at default is not defined by subsection (a), (b), or (c), it shall be determined by interpolation. To interpolate, locate the values between the nearest time before default and the nearest time after default for which a value is established by subsection (a), (b), or (c) of this Section. These are the established prior value and the established subsequent value.
e. When the value at default is determined by interpolation pursuant to subsection (d), the interpolation shall be done by use of the straight-line method unless the lease provides for use of the 125% declining balance method in which case that method shall be used.
f. When the straight-line method is used, the value at default shall be equal to the established prior value less the value of: the product of (1) the difference between the established prior value and the established subsequent value and (2) the number of months from the time of the established prior value to the time of default, divided by the number of months from the time of the established prior value to the time of the established subsequent value.
g. When the 125% declining balance method is used, the value of the leased property shall be determined for each month beginning with the first month after the established prior value and continuing through the month of default. The value in the first month after the established prior value is equal to the established prior value less 1.25 times the amount of decline in value for one month under the straight-line method. The value in any subsequent month shall be determined by multiplying the value for the prior month by the declining balance fraction. The declining bal-

ance fraction is equal to the value in the first month after the established prior value, divided by the established prior value.

SECTION 17. Lessor's right to take possession after default.

a. Upon default by a lessee, and compliance by the lessor with Section 15, the lessor is entitled to possession of the leased property. The lessor may take possession of the property without judicial process only if possession can be taken without trespass and without the use of force or other breach of the peace.
b. The lessor is liable to the lessee for any damages arising out of any repossession in violation of this Section.

SECTION 18. Supplementary General Principles of Law Applicable.

Unless displaced by the particular provisions of this Act, other principles of law relative to contracts and consumer protection shall apply to lease contracts.

SECTION 19. Penalties.

a. When a lease contract contains a provision inconsistent with the provisions of this Act, or made unenforceable by this Act, the lessee shall be entitled to recover an amount equal to the sum of:
1) actual damages sustained as a result of the violation;
2) 10% of the total amount of periodic payments under the lease contract, or $1,000, whichever amount is greater; and
3) the cost of the action, together with reasonable attorney's fees determined by the court.
b. If the lessor's interest in a lease is assigned, both the original lessor and the assignee shall be liable for damages and penalties provided by this Section.
c. Multiple violations of the provisions of this Act by the lessor that occur within a single lease contract shall constitute a single violation for purposes of this Section.

Amendment to Retail Installment Sales Act:

a. Any lease of goods which includes an option to purchase and in which the payments prior to the option to purchase are equal to, or more than, the cash price of the goods plus interest at

prevailing commercial rates for the term of the lease, whether or not the lessee is permitted to terminate the contract early without penalty, is a retail installment contract. For purposes of this act, a series of leases is a single lease if:

1) the leases are of the same goods and to the same lessee; and

2) the goods remain in the possession of the lessee.

b. An "option to purchase" may be either (1) a term of a lease, or (2) an understanding (created by advertising or any oral or written representations made by the lessor) between the parties to a lease, which provides that the lessee has the right to acquire ownership of the leased goods.

NOTE

JUDICIAL PROTECTION OF BALLOT-ACCESS RIGHTS: THIRD PARTIES NEED NOT APPLY

BRADLEY A. SMITH*

The states widely regulate third-party access to the general election ballot. One ubiquitous statutory regulation is the petition requirement, whereby a third party must gather a certain number of signatures in order to qualify for the ballot. Such statutes have been contested in the federal courts on constitutional grounds since 1968, when the Supreme Court struck down Ohio's ballot-access system in Williams v. Rhodes.

In this Note, Mr. Smith argues that subsequent decisions have failed to protect the constitutional right, recognized in Williams, *of individuals to associate for the advancement of political beliefs. Mr. Smith argues that this failure stems from a judicial refusal to recognize the onerous real-world burdens which ballot-access laws place upon third parties. Mr. Smith concludes that much ballot-access legislation is overly restrictive, and that the courts should apply true strict scrutiny to reform the two-party hegemony over ballot access.*

> All political ideas cannot and should not be channeled into the programs of our two major parties. History has amply proved the virtue of political activity by minority, dissident groups, which innumerable times have been in the vanguard of democratic thought and whose programs were ultimately accepted . . . . The absence of such voices would be a symptom of grave illness in our society.[1]

The last decade of the twentieth century has begun with a flurry of democratic activity in previously undemocratic states, from Eastern Europe to Mongolia, Nepal to Nicaragua. Efforts to conduct free elections in nations lacking democratic traditions have focused attention on the proper role of the state in regulating parties, campaigns, and access to the ballot.[2] These dem-

---

* J.D., Harvard Law School, 1990; Associate at Vorys, Sater, Seymour & Pease, Columbus, Ohio; Senior Policy Analyst, The Mackinac Center, Midland, Michigan; Policy Analyst, The Heartland Institute, Chicago, Illinois. The author thanks Professor Christopher Edley of Harvard Law School for his comments, and Mr. Richard Winger for his extensive assistance.

[1] Williams v. Rhodes, 393 U.S. 23, 39 (1968) (Douglas, J., concurring) (quoting Sweezy v. New Hampshire, 354 U.S. 234, 250–51 (1957) (opinion of Warren, C.J.)).

[2] *See, e.g., UN Fears for Latin Election*, Boston Globe, Feb. 6, 1990, at 12, col. 1 (UN concern over misuse of state property to support Sandinista campaign in Nicaragua); *Romanian Parties Will Join Coalition*, Boston Globe, Feb. 2, 1990, at 4, col. 5 (accusations that provisional government is manipulating scheduled elections and creating sham parties to mislead voters).

ocratic movements have sought to promote "multi-party"—not "two-party"—systems and elections.[3]

Yet in the United States the political system is under the firm control of two major parties. So powerful is this dominance that since 1928 only one third-party candidate for President has received more than seven percent of the vote.[4] One factor in this sustained dominance is state ballot-access legislation that is designed to present obstacles to third parties. This Note will argue that much of this legislation is unconstitutional, that the Supreme Court has failed to acknowledge this unconstitutionality because of its refusal to consider the real-world effects of ballot-access restrictions, and that stricter constitutional scrutiny is warranted.

Part I of the Note presents a brief history of third parties and of ballot-access laws. Part I also summarizes the content of ballot-access laws throughout the United States. Part II reviews the Supreme Court's ballot-access jurisprudence. Part III argues that the Court's approach is a failure because that approach is divorced from the realities of third-party activity in the United States. The Note concludes with a proposal for improving protection of the constitutional rights upon which ballot-access legislation infringes.[5]

---

[3] *See, e.g., Protest Stopped, 150 Seized in Nepal*, Boston Globe, Feb. 26, 1990, at 16, col. 6; *Whites Urged to Rise Against deKlerk*, Boston Globe, Feb. 6, 1990, at 11, col. 3; *Gorbachev Tells Party to Ease Grip*, Boston Globe, Feb. 6, 1990, at 1, col. 4.

[4] CONGRESSIONAL QUARTERLY, GUIDE TO U.S. ELECTIONS 349–66 (1985) [hereinafter U.S. ELECTIONS]. George Wallace received 13.5% of the vote in 1968. *See supra* notes 54–72 and text.

[5] For the purposes of this Note, "third party" and "minor party" are used interchangeably to designate all American political parties since 1832, other than the Republican and Democratic parties. However, the Republicans were considered a third party from their founding in July 1854 through the elections of 1856. Also, the Whigs were considered a major party from 1832 through the 1856 elections, when they were replaced by the Republicans.

The term "ballot-access laws" refers to state legislation that governs which candidates will appear on the general election ballot. This Note is not concerned with regulations governing the primary ballot except as they affect third-party access to the general election ballot. Also, the Note only addresses those laws limiting the ability of third parties to nominate a candidate, and does not discuss restrictions designed to limit particular individuals' ability to run without regard to their party label.

Most states have slightly different legislation governing independent candidates and third-party candidates. The Note attempts, therefore, to refer specifically to independent candidates when they are included in a bit of datum or covered by a referenced law. The problems of independent candidates are somewhat different from those faced by third parties, primarily in that third parties have a greater interest in long-term party building and incremental growth. *See* Hocker, *Legal Barriers to Third Parties*, 10 REV. L. & SOC. CHANGE 125 (1981). Generally, however, the constitutional issues raised by third-party and independent candidates are the same.

I. BALLOT ACCESS RESTRICTIONS: YESTERDAY AND TODAY

A. *Third Parties in American Politics*

Competitive third parties are "integral elements" of our political system.[6] Third parties serve a fundamental role in our "two-party system" as constructive channels through which those ignored by the major parties can voice dissent while maintaining allegiance to democratic norms.[7] Their existence as an alternative for dissatisfied voters provides a constant check on the major parties. "Their [the major parties'] performance is likely to be most satisfactory, therefore, when they live in danger of displacement by new parties."[8] In appealing to discontented voters, third parties have traditionally been an important source of policy innovations, including direct election of senators, women's suffrage, nomination through party primaries, the eight-hour work day, child labor laws, federal farm aid, and the graduated income tax.[9] Healthy third parties work against voter apathy by educating voters about neglected issues and providing an alternative for those who would otherwise stay at home on election day.[10]

Since modern American political parties first emerged in the 1830's, bursts of significant third-party activity have been more common than most people realize, occurring about every twenty years. Prior to the adoption of ballot-access laws, third parties regularly mounted serious challenges to the existing parties.

---

6 V.O. KEY, POLITICS, PARTIES, AND PRESSURE GROUPS 279 (5th ed. 1964). *But see* L. SABATO, THE PARTY'S JUST BEGUN: SHAPING POLITICAL PARTIES FOR AMERICA'S FUTURE (1988). Sabato concludes that the Republican and Democratic parties should be given "what we might term 'most favored nation' status in our laws." *Id.* at 179. Yet even Sabato admits that third parties play useful supporting roles in his vision of a two-party system. *Id.* at 40.

7 A. RANNEY & W. KENDALL, DEMOCRACY AND THE AMERICAN PARTY SYSTEM 457 (1956); W. GOODMAN, THE TWO PARTY SYSTEM IN THE UNITED STATES 50 (1960).

8 D. MAZMANIAN, THIRD PARTIES IN PRESIDENTIAL ELECTIONS 150 (1974). *See also* G. SARTORI, PARTIES AND PARTY SYSTEMS 25 (1976) (third parties "link people to a government"); V.O. KEY, *supra* note 6, at 258–59 (minor parties demonstrate the existence of voting blocks that can be wooed by a major party willing to take a stand on the issues raised by the minor parties).

9 S. Rosenstone, R. Behr & E. LAZARUS, THIRD PARTIES IN AMERICA 8 (1984) [hereinafter S. ROSENSTONE]; V.O. KEY, *supra* note 6, at 258. This role led John D. Hicks, former President of the American Academy of Historians, to declare that a vote for a third party is "probably the most powerful vote that has ever been cast." Hicks, *The Third Party Tradition in American Politics*, 20 MISS. VALLEY HIST. REV. 3, 27 (1933).

10 D. MAZMANIAN, *supra* note 8, at 149.

Although only the Republican Party succeeded in displacing a major party, historically third parties have elected large numbers of officials, presented viable alternatives to voters, and forced major changes in established party positions.

In the 1850's, the American, or Know-Nothing Party, competed with the Republicans to displace the decaying Whigs as a major party.[11] In the 1854 Massachusetts elections, the Know-Nothings won the Governor's office, the entire state senate, and all but two seats of the state house. The following year they captured legislative majorities in Rhode Island, New Hampshire, Connecticut, Maryland, and Kentucky. Their 1856 presidential candidate, former President Millard Fillmore, received 21.5% of the popular vote.[12]

In 1876 the Greenback Party, capitalizing on farm unrest, elected fourteen U.S. congressmen.[13] In 1892, the Populist Party polled over eight percent of the vote for President, won twenty-two electoral votes, and elected several congressmen, three governors, and hundreds of local officials.[14]

In the elections of 1912, third parties had their last major hurrah. Former President Theodore Roosevelt's Progressive Party actually ran ahead of the Republicans in the popular vote for President, and the Party elected fourteen congressmen.[15] With the "progressive" Woodrow Wilson in the White House, however, the Progressive vote fell off dramatically, assuring that the Republicans would remain the primary opposition party.[16]

Support for the Socialist Party peaked at six percent of the presidential vote in 1912. The Socialists elected a handful of congressmen, over 1200 local officials, and seventy-nine mayors.[17] In 1911 no fewer than thirty-three cities were under Socialist administration, including Milwaukee, Wisconsin; Berkeley, California; Butte, Montana; and Flint and Jackson, Michigan.[18]

After the elections of 1912, state legislatures began to write or revise ballot-access laws to hinder third-party activity. By

---

[11] *Id.* at 40–43.
[12] S. ROSENSTONE, *supra* note 9, at 56–58.
[13] V.O. KEY, *supra* note 6, at 256.
[14] *Id.* at 257.
[15] S. ROSENSTONE, *supra* note 9, at 86–87.
[16] *See* CONGRESSIONAL QUARTERLY, GUIDE TO THE PRESIDENCY 250 (1989).
[17] S. ROSENSTONE, *supra* note 9, at 89–90.
[18] D. SHANNON, THE SOCIALIST PARTY IN AMERICA 5 (1955).

1924 U.S. Senator Robert LaFollette, trying to launch a new Progressive Party, faced ballot-access laws that were "an almost insuperable obstacle to a new party."[19] Although third-party activity has continued to peak in roughly twenty-year cycles, no third party has been able to achieve electoral success even close to that routinely obtained by third parties prior to the First World War.[20]

Between 1896 and 1944, 128 congressmen were elected from third parties. No congressperson, and only one U.S. senator, has been elected by a third party since 1944.[21] Of 7461 state legislators serving in 1989, all but four independents were elected as Republicans or Democrats.[22] Of over 20,000 elections for state legislatures from 1982 through 1988, only three have been won by members of a party other than the Democrats or Republicans.[23]

The thorough electoral failure of third parties in the present day United States cannot be blamed entirely, or even primarily, on ballot-access laws. The complex nature of modern party organization and the formal and informal barriers otherwise present in the system effectively block broad third-party success. Nevertheless, strict ballot-access restrictions have helped the two major parties to achieve a vise-grip on American politics never before attainable. We turn now to the evolution of those restrictions.

---

[19] K. MacKay, The Progressive Movement of 1924, at 179 (1947).

[20] Major third-party efforts since the 1924 Progressives include the 1948 Progressive and States Rights Parties, the American Independent Party in 1968, and the Libertarian Party, New Alliance Party, and independent candidacy of John Anderson in the 1980's.

[21] *See* U.S. Elections, *supra* note 4, at 576–636, 833–1061. The U.S. senator was James Buckley, elected as a Conservative in New York in 1970. Buckley ran for reelection as a Republican. Seven independents have been elected to the House and two to the Senate since 1944. Both independent senators, Harry Byrd of Virginia and Strom Thurmond of South Carolina, were first elected as Democrats. *Id.* Bernard Sanders, a Vermont Socialist running as an independent, was elected to the House of Representatives in 1990. N.Y. Times, Nov. 12, 1990, at B9, col. 4.

[22] *See* Council of State Governments, State Elective Officials and the Legislatures, 1989-90, at v (1989). These figures do not include party affiliations for legislators in Nebraska, which holds nonpartisan elections.

[23] *See id.* and previous editions at vi (1987), vii (1985), vi (1983), which show two third-party legislators. The third is Nebraska State Senator Ernie Chambers. While Nebraska's legislative elections are technically nonpartisan, *see supra* note 22, Chambers is a member of the New Alliance Party. Formerly a Democrat, he was re-elected by write-in vote. Letter from Richard Winger, Field Representative, Coalition for Free and Open Elections, to Brad Smith (June 30, 1990) (on file at the Harv. J. on Legis.).

B. *American Ballot-Access Laws: 1888 to the Present*

The Constitution grants the states primary responsibility for the regulation of elections.[24] However, before the late nineteenth century, states did not regulate which parties and candidates would appear on the general election ballot because there were no official printed ballots prior to that time. Early elections were conducted by voice vote, or crude bean or corn ballots. As political parties emerged in the early nineteenth century, they began to print their own ballots, or "tickets," listing only their own candidates for office. These were distributed to as many voters as party resources allowed.[25]

The growth of corrupt political machines in the latter half of the nineteenth century led to widespread demands for new methods of electing public officials. Because the state had no control over the number of printed ballots, the unregulated system was ripe for ballot-box stuffing. In addition, each party typically printed ballots in the color of its choice, which made secret balloting all but impossible and allowed for regular episodes of bribery, coercion, and intimidation.[26]

The solution to these problems came in the form of the Australian ballot—the secret, uniform, government-printed ballot known to voters today. The Australian ballot swept across the country in the same wave of political reform that led to the direct election of senators, primary elections, the short ballot, and, ultimately, women's suffrage. The system was first used statewide in Massachusetts in 1888. By 1900, thirty-nine states had employed Australian ballots.[27]

The Australian ballot was praised as a device that would open up the two-party system to challenge by third parties. It was hoped that the secrecy of the ballot would not only prevent bribery and outright intimidation, but also the subtler sanctions of ridicule, dislike, and social or commercial injury.[28] As a result,

[24] U.S. CONST. art. II, § 1, cl. 2; art. I, § 4, cl. 1.

[25] B. ROBECK, 3 BALLOT ACCESS 5 (1978).

[26] The variety of 19th-century voting frauds is too extensive to list. Various examples of fraud and the general problems of voting in the pre-Australian ballot era can be found in a variety of sources, including B. ROBECK, *supra* note 25, at 5; D. MAZMANIAN, *supra* note 8, at 90; J. WIGMORE, AUSTRALIAN BALLOT SYSTEM 1–34 (1889); W. GOODMAN, *supra* note 7, at 437–38; Goldberg, *Election Fraud: An American Vice*, in ELECTIONS AMERICAN STYLE 182 (A.J. Reichley ed. 1987); S. ROSENSTONE, *supra* note 9, at 20.

[27] *See* A. LUDINGTON, AMERICAN BALLOT LAWS: 1888 - 1910, at 87 (1911).

[28] S. ROSENSTONE, *supra* note 9, at 25; J. WIGMORE, *supra* note 26, at 31.

the Australian ballot would break political machines and allow new political competitors to compete on more equal terms with established parties. In addition, early advocates of the Australian ballot argued that requiring candidates or parties to print their own ballots had effectively excluded all but the rich and established parties from the political system.[29] The Australian system, by contrast, would provide a "place on the ballot free to all nominees."[30]

Though it was intended to open up the political system, use of the Australian ballot required that the states adopt some mechanism for determining which candidates' names should appear on the official ballot. At a minimum, the state had to establish some cut-off date for nominations so that it could print ballots.[31]

Still, early laws were not restrictive, and there was a broad public consensus that ballot-access laws were not intended to be a substantive barrier to organized political parties appearing on the general election ballot. Wrote one influential commentator, "The real restrictions on the number of candidates will be found to be . . . public opinion and the interests of the aspirants."[32] He concluded that as few as two signatures on a nominating petition should be sufficient to secure a place on the ballot.[33]

However, "[w]ithin a very few years these laws underwent a number of changes that are difficult to justify as furthering the government's mandate to conduct efficient and honest elections."[34] The partisan political interests of legislators writing the ballot-access laws seem likely to have been the reason behind these changes. The first wave of restrictive laws came during

---

[29] W. Ivins, Electoral Reform and the History of the Yates-Saxton Bill, *quoted in* J. Wigmore, *supra* note 26, at 25–26.

[30] J. Wigmore, *supra* note 26, at 34.

[31] V.O. Key, *supra* note 6, at 640–41.

[32] J. Wigmore, *supra* note 26, at 53.

[33] *Id. See also* D. Mazmanian, *supra* note 8, at 91 ("The public, courts, and state officials apparently interpreted the mandate under the Australian ballot laws to be one of conducting expedient and honest elections—nothing more, nothing less."). The intended correction for voter confusion was widely thought to lie in reducing the number of offices voted on, not in restricting the number of candidates. *See* W. Goodman, *supra* note 7, at 448. Sixteen states originally adopted the Australian ballot with no ballot-access requirements on candidates: Alabama, Connecticut, Georgia, Kansas, Louisiana, Maryland, Massachusetts, Minnesota, Nebraska, North Carolina, Oklahoma, Pennsylvania, Rhode Island, South Dakota, Virginia, and Wyoming. *See* A. Ludington, *supra* note 27, at 12–93.

[34] D. Mazmanian, *supra* note 8, at 91.

the Wilson administration. After the 1912 Progressive Party candidacy of Theodore Roosevelt and the post-World War I "red scare," legislators began to exercise greater control over ballot access. By the mid-1920's, state laws governing access to the general election ballot were heavily weighted in favor of the existing major parties.[35]

Nevertheless, a second major wave of restrictive laws crested in the 1930's and 1940's, also due largely to the fear of communist parties. During this period, a number of states explicitly banned the Communist Party, while others continued to tighten theoretically "neutral" restrictions, primarily by requiring large numbers of signatures on nominating petitions.[36]

After a quiet period from the early 1950's through the middle 1960's, the evolution of ballot-access restrictions took a new twist in 1968. That year, the Supreme Court held that an Ohio statutory scheme that threatened to keep George Wallace's American Independent Party off the ballot was unconstitutional.[37] Although state court challenges to ballot-access laws had occurred within years of the adoption of the Australian ballot,[38] federal courts had previously stayed away from the issue,[39] and no significant, thematic body of law was developed prior to 1968.

In striking down the challenged Ohio statutes, the Supreme Court thrust the judiciary squarely onto the stage. The story of ballot-access laws since 1968 is largely one of court battles. However, before turning to that story, one preliminary remains: a description of the content of contemporary ballot-access legislation.

C. *State Ballot-Access Legislation*

Common tactics that states use to limit independent and third-party access to the ballot include loyalty oaths and outright bans

---

[35] *See* Hocker, *supra* note 5, at 126; D. MAZMANIAN, *supra* note 8, at 92.

[36] D. MAZMANIAN, *supra* note 8, at 92. *See also* Bone, *Small Political Parties: Casualties of War?*, 32 NAT'L MUN. REV. 524 (1943).

[37] *Williams*, 393 U.S. at 23, discussed *infra* notes 54–73 and text.

[38] *See, e.g.*, People *ex rel.* Hotchkiss v. Smith, 206 N.Y. 231, 99 N.E. 568 (1912) (finding unreasonable a requirement that more than 500 signatures appear on ballot); DeWalt v. Bartley, 146 Pa. 529, 24 A. 185 (1892) (finding unconstitutional a requirement that a party gather three percent of vote in previous election or file petition).

[39] *See* Snowden v. Hughes, 321 U.S. 1, 7 (1944) ("The right to become a candidate for state office, like the right to vote for the election of state officers, is a right or privilege of state citizenship . . . .") (citations omitted).

on subversive parties,[40] candidate filing fees,[41] and early filing deadlines.[42] However, the heart and soul of third-party ballot-access legislation is the petition drive, provided for in some form by every state except Mississippi.[43] Typically, a state's ballot-access statutes specify that those parties which received a certain percentage or number of votes in the last general election will be automatically placed on the ballot. Because most independent, new-party, and other third-party candidates do not usually meet this requirement, they must gather a pre-determined number of signatures on nominating petitions in order to be placed on the ballot. This number is usually expressed as a

---

[40] *See, e.g.*, GA. CODE ANN. § 21-2-2(20) (1987); ILL. ANN. STAT. ch. 46, para. 10-2 (Smith-Hurd Supp. 1989); KAN. STAT. ANN. § 25-117 (1986); MASS. GEN. LAWS ANN. ch. 50, § 1 (West Supp. 1989).

[41] *See, e.g.*, MD. ELEC. CODE ANN. art. 33, § 4A-6 (1976); MINN. STAT. § 204 B.11, subd. 1 (Supp. 1990); NEV. REV. STAT. § 293.193 (Supp. 1989); N.H. REV. STAT. ANN. § 655.19 (Supp. 1989); W. VA. CODE § 3-5-8 (1987).

[42] *See, e.g.*, ME. REV. STAT. ANN. tit. 21, § 494.9 (1983) (requires filing by April 1); MINN. STAT. § 204 B.09, subd. 1 (Supp. 1989) (requires filing between 56 and 70 days before major-party primaries).

[43] ALA. CODE § 17-16-2 (1987) (20% of vote in last election required for automatic ballot placement); ALASKA STAT. § 15.60.010(20) (1989) (3%); ARIZ. REV. STAT. ANN. § 16-804 (1984) (5%); ARK. STAT. ANN. § 7-1-101(1)(A) (Supp. 1989) (3%); CAL. ELEC. CODE § 6430(a) (West 1977) (2%); COLO. REV. STAT. § 1-1-104(18) (Supp. 1989) (10%); CONN. GEN. STAT. ANN. § 9-372(6) (West 1989) (1%); D.C. CODE ANN. § 1-1312 (1987) (7500 votes for District offices; for President, party must have won Presidency at least once since 1950); FLA. STAT. ANN. § 97.021(14) (West 1982) (5% of registered voters, regardless of the party's vote total in the last election); GA. CODE ANN. § 21-2-2(21) (1987) (20%); HAW. REV. STAT. § 11-61 (1988) (10% of total votes plus at least 10% in one-half of state legislative races); IDAHO CODE § 34-501(1)(b) (Supp. 1989) (3%); ILL. ANN. STAT. ch. 46, para. 10-2 (Smith-Hurd Supp. 1989) (5%); IND. CODE ANN. § 3-8-4-1 (Burns 1988) (2%); IOWA CODE § 43.2 (Supp. 1989) (2%); KAN. STAT. ANN. § 25-302(b) (1986) (1%); KY. REV. STAT. ANN. § 118.325(1) (Michie/Bobbs-Merrill Supp. 1989) (2%); LA. REV. STAT. ANN. § 18:441 (West 1979) (5%); ME. REV. STAT. ANN. tit. 21, § 321.1.C (1983) (5%); MD. ELEC. CODE ANN. art. 33, § 4C-1(a)(b) (Supp. 1989) (3%); MASS. GEN. LAWS ANN. ch. 53, § 1 (West 1975) (3%); MICH. COMP. LAWS § 168.685(6) (1989) (1%); MINN. STAT. § 204B.03 (Supp. 1989) (5%); MO. ANN. STAT. § 115.013(10) (Vernon Supp. 1990) (2%); MONT. CODE ANN. § 13-10-601(1) (1989) (5%); NEB. REV. STAT. § 32-521 (1988) (5%); NEV. REV. STAT. § 293.1715(2)(a) (Supp. 1989) (3%); N.H. REV. STAT. ANN. § 655 (1986) (3%); N.J. REV. STAT. § 19:5-1 (1989) (10%); N.M. STAT. ANN. § 1-1-9(B) (1985) (5%); N.Y. ELEC. LAW § 1-104(3) (McKinney 1978) (50,000 votes for Governor); N.C. GEN. STAT. § 163-96(a)(1) (1987) (10%); N.D. CENT. CODE § 16.1-11-30(3) (Supp. 1989) (5%); OHIO REV. CODE ANN. § 3517.01(A) (Anderson Supp. 1989) (5%); OKLA. STAT. ANN. tit. 26, § 1-109 (West 1976) (10%); OR. REV. STAT. § 249.732(2) (1983) (5%); PA. STAT. ANN. tit. 25, § 2831(A) (Purdon Supp. 1989) (2% of statewide total plus 2% in each of at least 10 counties); R.I. GEN. LAWS § 17-12.1-12(A) (Supp. 1989) (5%); S.D. CODIFIED LAWS ANN. § 12-1-3(3) (Supp. 1989) (10%); TENN. CODE ANN. § 2-13-201(b)(1)(A) (1985) (5% to 20%, depending on the office); TEX. ELEC. CODE ANN. § 181.005(b) (Vernon Supp. 1990) (5%); UTAH CODE ANN. § 20-3-2.5(1) (Supp. 1989) (2%); VT. STAT. ANN. tit. 17, § 2103(23) (1982) (5%); VA. CODE ANN. § 24.1-1(7) (1985) (10%); WASH. REV. CODE § 29.01.090 (Supp. 1989) (5%); W. VA. CODE § 3-1-8 (1987) (1%); WIS. STAT. § 5.62(1)(b) (Supp. 1989) (1%); WYO. STAT. § 22-1-102(g) (Supp. 1989) (10%).

percentage of registered voters or of the vote total in the last general election. Placing a third-party U.S. Senate candidate on the general election ballot in 1990 required anywhere from 200 signatures in New Jersey to 181,421 signatures in Florida.[44]

These requirements, it should be pointed out, are generally intended as barriers to third parties, not to individual candidates. Not only are the Republican and Democratic candidates granted automatic access to the general-election ballot, but states rarely require large numbers of signatures for a candidate to appear on a major-party primary ballot. For example, forty-nine states required a third-party candidate for U.S. Senate in 1988 to gather signatures, with an average signature requirement of 17,281, while only nineteen states required a major-party candidate to submit any signatures, with the average requirement being just 1243.[45] The petition drive, sometimes referred to as a preliminary showing of support, is the basic substantive obstacle to ballot access facing any third-party or independent candidate.

In addition to substantive signature requirements, many states have procedural restrictions on the gathering of signatures.[46] Thirteen states require persons signing a petition to swear that they belong to the party, will vote for the candidates listed, or will otherwise support the nominee.[47] Other states hamper petitioning by requiring arcane information to be listed on the petition. For example, Texas and South Carolina require voter affidavit numbers on the petition.[48] As voters rarely know this

---

[44] Ballot Access News, Feb. 12, 1990, at 5, col. 1. Ballot Access News is published by Richard Winger, Field Representative for the Coalition for Free and Open Elections, San Francisco, Cal. It is an invaluable resource for anyone interested in monitoring ballot issues and the regulation and electoral success of third parties generally.

[45] H.R. 1582, 101st Cong., 1st Sess. § 2(a)(7) (1989) (congressional findings included in Fair Election Act).

[46] The following procedural restrictions are surveyed in R. Winger, Why HR 2320 Is Needed 5–6 (unpublished manuscript on file at the HARV. J. ON LEGIS.) [hereinafter Winger Manuscript].

[47] *Id.* at 6. *See* CAL. ELEC. CODE § 6430 (West 1977); DEL. CODE ANN. tit. 15, § 3001 (1981); HAW. REV. STAT. § 11-62(a)(2) (1988); ILL. ANN. STAT. ch. 46, para. 10-2 (Smith-Hurd Supp. 1989); IND. CODE ANN. § 3-8-6-5(5) (Burns Supp. 1989); MD. ELEC. CODE ANN. art. 33, § 4B-1(a) (1976); N.J. REV. STAT. § 19:13-4 (1989); N.Y. ELEC. LAW § 6-140 (McKinney 1978); N.C. GEN. STAT. § 163-96(b) (1987); OHIO REV. CODE ANN. § 3517.011 (Anderson Supp. 1989) (petition required by Secretary of State regulations); OR. REV. STAT. § 249.732(1) (1983); UTAH CODE ANN. § 20-3-2.5(2)(b)(i) (Supp. 1989); W. VA. CODE § 3-5-23(d) (1987). West Virginia's law was found unconstitutional in Socialist Workers Party v. Hechler, 890 F.2d 1303 (4th Cir. 1989).

[48] Winger Manuscript, *supra* note 46, at 6. *See* S.C. CODE ANN. § 7-11-80(3)(c) (Law. Co-op. Supp. 1989); TEX. ELEC. CODE ANN. § 141.063(2)(B) (Vernon 1986). A voter affidavit number is the number given to a voter when he signs an affidavit to certify that he is eligible to register to vote in that state. The Texas statute was held unconstitutional in Pilcher v. Rains, 853 F.2d 334 (5th Cir. 1988). However, because *Pilcher* only

trivia, petitioners must expend considerable effort looking up the information.[49] Similarly, Kentucky and Delaware require signers to provide their social security numbers.[50] At least six states prohibit persons from signing petitions if they vote in a party primary.[51] Eight states require candidates wishing to appear on a statewide ballot to obtain a certain number of signatures from different congressional or legislative districts.[52] Seventeen states limit the time in which petition signatures may be gathered, often to quite brief periods.[53]

Such additional restrictions on signature gathering are an important aspect of ballot-access legislation. However, the focus of constitutional challenges to ballot-access laws has been on whether it is justifiable to require large numbers of signatures to be gathered in the first place. Accordingly, the following sections will concentrate on the constitutionality of the core ballot-access petition and retention requirements. This Note will discuss other restrictions only as they relate to, or shed light on, this fundamental issue.

---

specifically addressed third parties, the state has continued to enforce the provision against independent candidates. In September 1990, an independent candidate was kept off the Texas ballot for failure to include affidavit numbers on her petitions. Ballot Access News, Oct. 9, 1990, at 5, col. 2.

[49] Winger Manuscript, *supra* note 46, at 6.

[50] *Id. See* DEL. CODE ANN. tit. 15, § 3002(C)(2) (1981); KY. REV. STAT. ANN. § 118.315(2) (Michie/Bobbs-Merrill Supp. 1990). Demanding social security numbers violates the Federal Privacy Act of 1974, 5 U.S.C. § 552(a) (1988 & Supp. 1990). In 1983 Kentucky's Attorney General advised the legislature of the problem, KY OAG 83-437 (1983), but the legislature has not changed the law. Winger Manuscript, *supra* note 46, at 6.

[51] Winger Manuscript, *supra* note 46, at 6. *See* ARIZ. REV. STAT. ANN. § 16-341(C) (Supp. 1986); ILL. ANN. STAT. ch. 46, para. 10-4 (Smith-Hurd Supp. 1990); NEB. REV. STAT. § 32-504(3)(e) (1989); N.Y. ELEC. LAW § 6-138(1) (McKinney 1978); TEX. ELEC. CODE ANN. § 181.006(g) (Vernon Supp. 1990); W. VA. CODE § 3-5-23(c) (1987) (petitioners must advise signers that they may not vote in the primary, which is held later).

[52] Winger Manuscript, *supra* note 46, at 5. *See* MICH. COMP. LAWS § 168.685(1) (1989); MO. ANN. STAT. § 115.315(4) (Vernon Supp. 1990); MONT. CODE ANN. § 13-10-601(2) (1989); NEB. REV. STAT. § 32-526(1) (1988); N.H. REV. STAT. ANN. § 655.42(1) (1986); N.Y. ELEC. LAW § 6-142.1 (McKinney 1978); N.C. GEN. STAT. § 163-96(a)(2) (1987); VA. CODE ANN. § 24.1-168 (Supp. 1989).

[53] ARIZ. REV. STAT. ANN. §§ 16-341C & G (Supp. 1986) (10 days); ARK. STAT. ANN. § 7-7-103(c)(3) (1990) (60 days); COLO. REV. STAT. §§ 1-4-801(e) & (h) (Supp. 1989); FLA. STAT. ANN. § 99.0955(1) (West 1982); GA. CODE ANN. § 21-2-170(e) (1987); ILL. ANN. STAT. ch. 46, para. 10-4 (Smith-Hurd Supp. 1989) (90 days); MD. ELEC. CODE ANN. art. 33, § 4B-1(a) (1976); MICH. COMP. LAWS § 168.685(1) (1989); MINN. STAT. § 204B.08 subd. 1 (Supp. 1989) (14 days); N.Y. ELEC. LAW § 6-138(4) (McKinney 1978) (6 weeks); OKLA. STAT. ANN. tit. 26, § 1-108(2) (West Supp. 1990); PA. STAT. ANN. tit. 25, § 2913 (Purdon Supp. 1989); R.I. GEN. LAW § 17-14-1 (Supp. 1989); TEX. ELEC. CODE ANN. § 181.006 (Vernon Supp. 1990); VA. CODE ANN. § 24.1-168 (Supp. 1989); WIS. STAT. § 8.20(8)(a) (1986); WYO. STAT. § 22-4-201 (Supp. 1989).

II. Supreme Court Analysis of Ballot-Access Rights

A. *Framing the Issue:* Williams v. Rhodes *and* Jenness v. Fortson

The Supreme Court first addressed the constitutional status of state ballot-access laws in *Williams v. Rhodes*.[54] Though *Williams* was decided during the heat of the 1968 presidential campaign, the history of the suit begins twenty years earlier, in the aftermath of the 1948 presidential election.

In that election, the Progressive Party of Henry Wallace received 1.3% of the presidential vote in Ohio, an amount sufficient to prevent either major party from attaining fifty percent of the vote.[55] The Ohio legislature responded by adopting a highly restrictive ballot-access system that essentially limited the ballot to the Republican and Democratic parties.[56] This maze of regulations required a new party to obtain petition signatures equal to fifteen percent of the vote cast in the last gubernatorial race, file such petitions approximately nine months before election day, and establish and maintain an extremely detailed, elaborate party structure. The system also banned all write-in voting.[57]

In January 1968, supporters of George Wallace organized the Ohio American Independent Party and conducted a six-month petition-signature drive. The Party eventually submitted over 450,000 signatures to the Ohio Secretary of State. This number exceeded the 433,100 required by state law, but the signatures were filed substantially after the statutory deadline. Wallace was denied a place on the ballot.

---

[54] 393 U.S. 23 (1968).

[55] Truman carried the state for the Democrats with 49.5% of the vote, a 7107-vote margin over Republican candidate Dewey. Wallace attracted 37,596 votes. R. Scammon, America at the Polls: The Vote for President 1920–1964, at 15 (1965). It is widely believed that Wallace nearly threw the state to Dewey by taking more votes from Truman than Dewey. *See* Smithsonian Institution Press, 'If Elected . . .' Unsuccessful Candidates for the Presidency, 1796–1968 (1972). However, Truman would have won the election even if Dewey had carried Ohio. *See* R. Scammon, *supra*, at 15–16.

[56] Between 1950 and 1966, no third-party candidates appeared on any Ohio ballot. *See* U.S. Elections, *supra* note 4, at 358–61. Including Wallace's Progressives, third parties had qualified for the state ballot a total of nine times in the nine elections between 1932 and 1948. *Id.* at 353–57, 371, 372.

[57] *Williams*, 393 U.S. at 24–27.

The Party filed suit in U.S. District Court, claiming that the Ohio statutes violated the fourteenth amendment's equal protection clause. A three-judge panel ruled that the state must provide a write-in option, but refused to place the Party on the ballot.[58] The Party appealed to the Supreme Court; Justice Stewart granted a temporary injunction placing the Party on the ballot pending appeal; and the Court set the case for oral argument. Because of the time constraints of the campaign, the Court took just seven days to render a decision.[59]

In an opinion written by Justice Black,[60] the Court held that the Ohio law violated the first and fourteenth amendments by burdening "two different, although overlapping, kinds of rights."[61] The first of these was "the right of individuals to associate for the advancement of political beliefs."[62] The second was "the right of qualified voters . . . to cast their votes effectively."[63] Though Ohio did not prohibit the formation of political parties, the right of association could be rendered meaningless by prohibiting parties from appearing on the ballot. Likewise, the right to vote would lose its meaning if voters were limited to just one or two government-approved parties.[64]

Having found an infringement of "fundamental rights," the Court held that only a compelling state interest could justify Ohio's regulations.[65] None of the four interests advanced by Ohio qualified as compelling state interests.

First, the state claimed to have an interest in promoting compromise and stability through a two-party system. The Court found this insufficient to justify the burdens imposed on fundamental rights because the system promoted not a generic "two-party" system, but two specific parties—Republicans and Democrats.[66]

---

[58] Williams v. Rhodes, 290 F. Supp. 983 (S.D. Ohio 1968).

[59] *Williams*, 393 U.S. at 63 (Warren, C.J., dissenting).

[60] Justices Douglas, Brennan, Marshall, and Fortas joined Justice Black's opinion for the Court. Justice Douglas also concurred separately. Justice Harlan concurred in the result, arguing that the issue was one of due process, not equal protection. *Id.* at 43 (Harlan, J., concurring). Chief Justice Warren and Justices Stewart and White wrote separate dissents.

[61] *Id.* at 30.

[62] *Id.*

[63] *Id.*

[64] *Id.* at 31.

[65] *Id.*

[66] *Id.* at 31–32.

Second, Ohio asserted its interest in assuring majority rather than plurality election winners. This was rejected by the Court as an interest that could not be served without placing overly severe impositions on minor parties.[67]

Third, the state claimed an interest in preventing the factionalism that results when a large number of candidates enter an election. The Court found that though this was a desirable state interest, it could not be achieved by the Ohio system. The early filing deadlines either deprived factions of the opportunity to field an alternative candidate, or forced them to organize at a very early date, thus increasing, rather than decreasing, factionalism.[68]

Finally, the state asserted an interest in preventing voter confusion by reducing the number of candidates on the ballot. This argument failed as contrary to actual experience in Ohio, because even when only one percent of the electorate's signatures had been required to place a party on the ballot, few parties attempted to qualify for ballot positions. The state's claim of voter confusion, therefore, was "no more than 'theoretically imaginable.'"[69]

The Court did not hold any particular provision of the Ohio law unconstitutional. Instead, the Court struck down "the totality of the Ohio restrictive laws taken as a whole,"[70] and ordered that the Ohio American Independent Party remain on the ballot.

Both logic and politics indicate that the Court reached the right result in *Williams*. Governor Wallace's supporters had submitted signatures of over fifteen percent of the number of voters in Ohio's last gubernatorial election. The case was heard near the peak of Wallace's popularity. His support in the most recent Gallup Poll was above twenty percent and rising, and he seemed certain to appear on all other state ballots and to win at least some electoral votes.[71] It was widely thought that he might even force the election into the House of Representatives. To

---

[67] *Id.* at 32.
[68] *Id.* at 32–33.
[69] *Id.* at 33 (quoting Mine Workers v. Illinois Bar Ass'n, 389 U.S. 217, 224 (1967)).
[70] *Id.* at 34.
[71] *Williams* was argued on October 7, 1968, and decided on October 15, 1968. The last Gallup Poll published before the Court issued its decision, on September 29, 1968, showed Wallace with 21% of the probable vote, and gaining ground on Democrat Hubert Humphrey, who was at 28% and falling. G. GALLUP, 3 THE GALLUP POLL: PUBLIC OPINION 1935–1971, at 2162 (1972).

have denied a ballot position to such a formidable candidate would have raised serious questions about whether American democracy was truly representative. A denial may also have raised discontent to the danger point among Wallace voters, who already felt alienated from and ignored by the political system.[72]

The language of *Williams v. Rhodes* was sufficiently broad to raise doubts as to whether any state's ballot-access law could pass constitutional muster.[73] Judging from the ballot-access decisions that followed *Williams*, however, the Court may have gone further than it intended. The history of ballot-access adjudication in the past twenty years is one of retreat from the broad implications of *Williams*.

This retreat began with the 1971 decision in *Jenness v. Fortson*.[74] At issue in *Jenness* was a Georgia election statute that granted automatic ballot status only to those parties that received twenty percent or more of the vote in the previous gubernatorial or presidential election. The statute required any other party or candidate to submit a nominating petition signed by at least five percent of the state's registered voters. Parties were restricted to gathering signatures over a 180-day period ending on the second Wednesday in June.[75]

The Court upheld the Georgia statute on the ground that it did not "operate to freeze the status quo," but rather offered a realistic possibility for third-party and independent candidates to obtain a place on the ballot.[76] The difference between the Court's approach in *Jenness* and its approach in *Williams* is striking.[77] Whereas *Williams* focused on infringements of fundamental rights, *Jenness* scarcely even considered such rights. Instead, the Court simply compared the Ohio and Georgia regulatory systems in detail and concluded that the Georgia scheme had "insulated not a single potential voter from the appeal of

---

[72] *See* Pettigrew, Riley, & Vanneman, *George Wallace's Constituents*, 5 PSYCHOLOGY TODAY 47 (Feb. 1972).

[73] *See The Supreme Court, 1968 Term*, 83 HARV. L. REV. 7, 96 (1969).

[74] 403 U.S. 431 (1971).

[75] *Id.* at 432.

[76] *Id.* at 438.

[77] *See* L. TRIBE, AMERICAN CONSTITUTIONAL LAW § 13-20, at 1106 (2d ed. 1988). Neither Court specifically mentioned a standard of review, but the *Williams* Court's emphasis on "compelling interests," 393 U.S. at 31, and quick dismissal of state claims indicate strict scrutiny. *See* L. TRIBE, *supra*, at 1102. *Jenness* implicitly uses only minimal scrutiny. *Id.* at 1105.

new political voices."[78] Among the decisive differences identified by the Court were that the Georgia system imposed no restrictions on write-in votes; it provided for independent candidacies; its filing deadline was not unreasonably early; and it did not require a third party to establish elaborate political machinery.[79]

In contrast to the *Williams* opinion, which had brusquely discounted Ohio's asserted "compelling state interests," the *Jenness* Court never seriously considered the possibility that such interests might not be valid. Justice Stewart, writing for the Court, devoted a single sentence to the issue: "There is surely an important state interest . . . the interest, if no other, in avoiding confusion, deception, and even frustration of the democratic process."[80] All of the contested Georgia statutory provisions were upheld, following the "totality" approach adopted in *Williams*.

The *Jenness* decision relied on a fundamentally flawed assessment of the actual effects of the Georgia ballot-access provisions. First, the Court failed to recognize the magnitude of the burden imposed by Georgia's five-percent signature requirement.[81] On its face, Georgia's five-percent requirement appears to be less than the fifteen percent required by the Ohio statute struck down in *Williams*. The Ohio law, however, was based on fifteen percent of the ballots cast in the last election, while Georgia's was based on the number of registered voters—a much larger base. Thus, the Georgia requirement was closer to Ohio's than the Court realized.[82] Moreover, although Justice Stewart acknowledged that "[t]he 5% figure is . . . somewhat higher than the percentage of support required to be shown in many States as a condition for ballot position,"[83] he failed to

---

[78] 403 U.S. at 442.

[79] *Id.* at 438.

[80] *Id.* at 442.

[81] Although *Williams* did not address the validity of Ohio's 15% signature requirement taken alone, this high percentage weighed heavily in the Court's decision. The *Williams* opinion noted that the 15% requirement was far higher than that required in most other states. *Williams*, 393 U.S. at 33 n.9. Justice Harlan, concurring in the *Williams* result, would have found the 15% requirement unconstitutional even in the absence of the other challenged provisions. *Id.* at 46 (Harlan, J., concurring).

[82] In at least one Georgia election since the *Jenness* decision, a new party wishing to appear on the general election ballot was required to submit signatures in excess of 15% of the ballots eventually cast. In 1978, third-party or independent candidates for governor of Georgia were required to submit 104,514 signatures, or 15.8% of the ballots cast. Winger Manuscript, *supra* note 46, at 3.

[83] 403 U.S. at 442.

note the extent of the disparity. Georgia's five-percent requirement was at least five times that of forty-two states, and at least fifty times that of sixteen states.[84]

Second, the *Jenness* Court did not address the impact of Georgia's ballot-retention requirements.[85] Once qualified in Ohio, a party could remain qualified by receiving just ten percent of the vote.[86] Georgia, however, required a party to receive twenty percent of the vote to retain automatic ballot status. Yet no third-party or independent candidate for governor, and only one for President, has received twenty percent of the Georgia vote since 1912.[87] Thus, the Georgia statute would require even a reasonably successful third party—one capable of receiving as much as fifteen to twenty percent of the vote—to continue to devote scarce resources to petition drives while Democrats and Republicans automatically received ballot status. Such a requirement is a significant barrier to parties "working to increase their strength from year to year," a concern explicitly voiced by the *Williams* Court in discounting Ohio's alleged interest in maintaining its ballot-access laws.[88]

Third, the Court ignored the fact that, by requiring a separate petition for each candidate, the Georgia law could have a more restrictive impact than the Ohio statute on a party hoping to nominate a full slate of candidates. In Ohio a new party could qualify all of its candidates for office with a single petition.[89] By contrast, a party in Georgia wishing to nominate the full complement of ten candidates for statewide office, plus candidates for the U.S. Senate and House, the state legislature, and a county office, would need to ask persons approached to sign at least fourteen different petitions.[90] Generally, petitioning is accomplished by interrupting passers-by on the street, few of whom are willing to devote more than a few moments to the petitioner. Thus, the system made gathering signatures in Georgia much more difficult than in Ohio.

---

[84] *See Williams*, 393 U.S. at 47 n.10 (Harlan, J., concurring).

[85] Although both *Williams* and *Jenness* involved challenges to the petition process, and not the retention question, the retention requirement should be considered an important element of the "totality" approach endorsed in both *Williams* and *Jenness*. *See id.* at 34; *Jenness*, 403 U.S. at 437.

[86] *Williams*, 393 U.S. at 26.

[87] U.S. ELECTIONS, *supra* note 4, at 497.

[88] 393 U.S. at 32.

[89] *See* Winger Manuscript, *supra* note 46, at 3.

[90] *Id.*

Finally, the *Jenness* Court overstated the historical ability of third parties to gain access to the Georgia ballot. The Court emphasized that, while the Ohio law had foreclosed any third-party competition, "[t]he open quality of the Georgia system is far from merely theoretical."[91] In support of this characterization, the Court noted that the petition procedure had been used once in 1966 and again in 1968. Prior to 1966, however, no candidate had ever successfully used the petition procedure to appear on Georgia's statewide ballot.[92] The candidate who qualified in 1966 was not a minor party or independent candidate, but the Republican nominee for Governor.[93] Thus, the differences between the Ohio and Georgia systems, in "totality," were not nearly so great as the Court's opinion in *Jenness* would make it seem.

The failure of the Court to assess accurately the restrictive effects of the Georgia ballot-access statute was not the only deficiency in the *Jenness* decision. The Court's conceptual analysis was also deeply flawed, and departed sharply from that of *Williams*. The *Jenness* Court argued that it was not inherently more burdensome to require a candidate to gather signatures than to win a party primary. Consequently, third-party candidates did not have any more of an equal protection claim than primary losers.[94] This approach confuses the right of an individual to appear on the general election ballot as a party's nominee with the right of individuals to form a party, nominate, and vote for a candidate. This distinction is at the core of both the *Williams* decision and the *Jenness* complaint.

Williams held that the right to organize a political party necessarily implies the ability of parties to place their candidates before the electorate.[95] Nevertheless, "these associational rights do not seem to require that any *particular* individual serve as [a party's] candidate."[96] Thus, a candidate who fails to gain the

---

[91] *Jenness*, 403 U.S. at 439.

[92] *See* Winger Manuscript, *supra* note 46, at 4.

[93] The 1968 qualifier was George Wallace. U.S. ELECTIONS, *supra* note 4, at 121.

[94] 403 U.S. at 440.

[95] 393 U.S. at 31.

[96] L. TRIBE, *supra* note 77, § 13-19, at 1098 n.5 (emphasis in original). In fact, many states, including Georgia, restrict the ability of a candidate to appear on the general election ballot after unsuccessfully seeking another party's nomination. Georgia's statute requires an independent or third-party candidate to file prior to the major-party primaries. GA. CODE ANN. § 21-2-132 (Supp. 1990). Other states specifically prohibit an individual from pursuing office as an independent or a member of a party if she was affiliated with another party within a specified period of time prior to the election. *See*,

nomination of either a major or minor party, whether by primary or other means of selection, is not entitled to appear on the ballot under that party's label. The *Jenness* Court, however, equated parties unable to meet state-imposed substantive showing of support requirements with primary losers. In essence, the Court jumped from the premise that an individual has no right to be the nominee of a particular party to the conclusion that a party has no right to place a nominee before the voters. Such a conclusion not only contradicts the rights stressed in *Williams*, but also serves to limit the exercise of these rights to members of the two major parties which have automatic ballot access.

The *Jenness* Court attempted to counter the argument that third parties were treated unfairly. It argued that the right to association in a party other than the two major parties was not abridged by Georgia because major parties were required to maintain an elaborate party structure to assure an automatic place on the ballot, while minor parties were not.[97] This argument ignores the "totality" approach the Court purported to be using, implying that not imposing these burdens on third parties meant the law was per se constitutional. But even if a third party did maintain such an organizational structure, it would not be placed on the ballot unless it received twenty percent of the vote in the last election or gathered signatures. It was this latter criterion, not organizational structure, that was specifically under attack.

The most curious thing about the *Jenness* decision is the Court's turnabout from *Williams*. While the Georgia law was not as restrictive as the statute struck down in *Williams*, the difference was not so great as the *Jenness* Court tried to make it appear. Furthermore, because *Jenness* found compelling those same state interests that *Williams* specifically rejected, this element of the decision approached outright reversal. Given the expansive holding of *Williams*, it is surprising that five Justices who had voted in *Williams* to strike down the Ohio law voted in *Jenness* to uphold the Georgia law.[98]

---

e.g., ARK. STAT. ANN. § 7-7-103(f) (Supp. 1989); CAL. ELEC. CODE ANN. § 6801 (West 1977). These "sore loser" laws withstood constitutional challenge in Storer v. Brown, 415 U.S. 724 (1974).

97 403 U.S. at 441.

98 Justices Douglas, Brennan, and Marshall joined the majority in *Jenness*. Justice Harlan concurred in the result in both cases. Justice Black, who wrote the *Williams* opinion, concurred in the *Jenness* result without an opinion.

One reason for the shift may have been the Court's experience with *Rockefeller v. Socialist Workers Party*.[99] In June 1970 a special three-judge district court declared unconstitutional a New York requirement that ballot-access petitions contain fifty signatures from each county in the state. The Supreme Court summarily affirmed. With the county distribution requirement eliminated, established candidates formed dummy "parties," which "nominated" them a second time and circulated ballot petitions. This enabled the candidates to appear on the ballot on multiple lines.[100] Nelson Rockefeller created the "Civil Service Independents Party," James Buckley created the "Independent Alliance Party," and Richard Ottinger, already on the ballot as both the Democratic and Liberal nominee, created the "Conservation Party." None of these parties existed apart from the candidates' campaign organizations. With seven "real" parties already on the ballot, the three dummy parties brought to ten the number of "parties" qualified for ballot status, and New York's mechanical voting machines had room for just nine. It appeared that the state's election machinery would be disrupted on the eve of the election. The problem was solved when the "Conservation Party" was removed from the ballot because its name was too close to that of the established Conservative Party. *Jenness* was the next ballot-access case to reach the Court, and the Court may have been influenced by a desire to avoid such crises in the future.[101]

Whatever the reasons behind it, *Jenness v. Fortson* was a disaster for third-party and independent candidates. Whereas *Williams* had left unclear which, if any laws, could withstand constitutional scrutiny, *Jenness* set virtually no upper limit as to how significant a showing of support a state could require before granting access to the ballot. So long as a law was flexible enough to allow an occasional third-party or independent candidate to qualify for the ballot, it seemed to satisfy the *Jenness* standard of openness. Whereas *Williams* had applied a rigorous strict scrutiny standard, *Jenness* applied minimal scrutiny.[102]

---

[99] 314 F. Supp. 984 (S.D.N.Y.), *aff'd mem.*, 400 U.S. 806 (1970). The discussion of the *Rockefeller* decision below follows the account set forth in Winger Manuscript, *supra* note 46, at 4.

[100] Most states force a candidate to choose just one party label under which to appear on the ballot. *See, e.g.*, IND. CODE ANN. § 3-8-7-21 (Burns 1988); KAN. STAT. ANN. § 25-306 (1986); WISC. STAT. § 8.03(1) (1986).

[101] Winger Manuscript, *supra* note 46, at 4.

[102] *See* L. TRIBE, *supra* note 77, § 13-20, at 1105.

Jenness appears to have been a catalyst, if not a cause, of renewed state efforts to restrict ballot access. In the first fifteen years after the *Jenness* decision, seventeen states raised their numerical requirements for a third party to qualify by petition—more than had raised them in the preceeding thirty years.[103]

B. *Post-*Jenness *Development of the Law*[104]

In the years since *Jenness v. Fortson*, the Supreme Court's ballot-access jurisprudence has shown an erratic but unmistakable trend toward a narrower definition of infringed rights, greater deference to state interests, and a less rigorous standard of review.[105] This trend began with a pair of cases decided on the same day in 1974, *Storer v. Brown*[106] and *American Party v. White*.[107]

Storer addressed provisions of the California Elections Code. The Code required that an independent candidate for office be free from any party affiliation for at least one year prior to the primary election for the office sought; file nominating petitions with signatures equal to at least five percent of the vote cast in the last general election; and gather such signatures within a twenty-four day period immediately following the primary. The statute also limited those eligible to sign the petitions to those who had not voted in the primary.[108]

Although the Court perfunctorily acknowledged that the California provisions burdened first and fourteenth amendment rights, it primarily emphasized the substantial state interests underlying the restrictions. These interests included political stability and compromise, an understandable ballot, assurance of majority winners, and prevention of "clogging of election

[103] Winger Manuscript, *supra* note 46, at 3. The 17 states were Alabama, Arizona, Arkansas, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maine, New Mexico, New York, North Carolina, North Dakota, Oklahoma, Pennsylvania, Rhode Island, and Virginia. *Id.*

[104] A particularly illuminating analysis of the post-*Jenness* development of the law is provided in L. TRIBE, *supra* note 77, § 13-20, at 1106. The discussion that follows benefits from Tribe's account.

[105] *But see id.* at 1110 ("It is still too early to predict whether these cases portend a general retreat from the rigid, two-tiered standard of equal protection . . . .").

[106] 415 U.S. 724 (1974).

[107] 415 U.S. 767 (1974), *reh'g denied*, 417 U.S. 926 (1974).

[108] 415 U.S. at 727.

machinery."[109] Noting that the provisions did not make it "virtually impossible" for new candidates and parties to appear on the ballot,[110] the Court found that the disaffiliation test and restrictions on petition signing by primary voters were justified by the state interest involved. The Court, however, chose to remand the petition issue, ordering the district court to determine whether the signature requirement and time limit for collecting signatures were reasonable within the context of California politics[111]—the first time the Court had suggested that a requirement might pass constitutional muster in one state but not another.

The *Storer* decision is significant for two reasons. First, although the Court found a compelling state interest in the disaffiliation provisions, it did not inquire into whether less drastic means were available to the state to serve that interest. Implicit in this approach is a standard of review significantly less demanding than traditional strict scrutiny.[112]

Second, the *Storer* decision indicated that the five-percent requirement upheld in *Jenness* was not the upper limit on the showing of support that might be found reasonable.[113] The Court recognized that if primary voters were not allowed to sign an independent candidate's nominating petition, the effect could be to require the candidate to gather signatures of far more than five percent of those voters eligible to sign. The Court, however, left it to the district court to determine if the effect of this provision would make the percentage of signatures unduly burdensome. By contrast, the dissent, noting that the actual percentage of signatures required was 9.5% of those voters eligible to sign, argued that such a high percentage served "no compelling state interest."[114]

---

[109] *Id.* at 728–32. The Court cited *Williams* for the proposition that assuring majority winners was a substantial state interest. *Id.* at 729. This is a misreading of *Williams*, for while the *Williams* Court recognized that interest, it found that the interest could not be enforced without violating the rights of parties which were "working to increase their strength from year to year." 393 U.S. at 32.

[110] 415 U.S. at 728 (quoting *Williams*, 393 U.S. at 25).

[111] *Id.* at 737–41.

[112] *Id.* at 760–61 (Brennan, J., dissenting). *See also* L. TRIBE, *supra* note 77, § 13-20, at 1107. An alternative analysis is that the Court was applying traditional strict scrutiny, but had explicitly identified a "compelling state interest" for the first time. *See* Jardine, *Ballot Access Rights: The Constitutional Status of the Right to Run for Office*, 1974 UTAH L. REV. 290, 301 (1974).

[113] Jardine, *supra* note 112, at 320.

[114] 415 U.S. at 764 (Brennan, J., dissenting). The 9.5% figure is nearly double the percentage of signatures required under the Georgia statute upheld in *Jenness*. Thus,

In *American Party v. White*, the Court upheld Texas' ballot-access system for reasons similar to those invoked in *Storer*. As in *Storer*, the Court found compelling state interests, but did not engage in a least restrictive alternative analysis.[115] The Texas statute required a third party to hold precinct nomination conventions and submit signatures equal to at least one percent of the vote in the most recent gubernatorial election. The parties were given just fifty-five days to gather signatures, and voters could not sign more than one petition. Primary voters were ineligible to sign. In upholding the Texas statute, the Court emphasized that the law did not freeze the status quo, and that two of the plaintiffs had, in fact, previously met the requirements.[116]

The Court did not hear another case challenging signature requirements until *Illinois State Board of Elections v. Socialist Workers Party*[117] in 1979. In *Illinois State Board*, the Court struck down an Illinois statute which had the effect of requiring a political party to file over 63,000 signatures to appear on the ballot in Chicago, but just 25,000 signatures to appear on a statewide ballot. The Court made no effort to decide whether the Chicago requirement, by itself, exceeded the showing of support that a state could require before granting ballot access. Instead, the Court used the traditional strict scrutiny test of requiring the state to use the "least restrictive means."[118] Because just 25,000 signatures was a sufficient showing of support to serve the state's interests regarding the larger, statewide electorate, requiring more for a local election was clearly not the least drastic means available to the state.[119]

Four years later, however, the Court added to the confusion over the proper standard of review by failing to apply strict scrutiny in the 1983 case of *Anderson v. Celebrezze*.[120] In *Anderson*, the Court struck down an Ohio law requiring independent candidates for President to file seventy-five days before the party primaries, more than seven months prior to the general

---

the *Storer* decision heightened the baseline level of acceptability in state petition requirements.

[115] *See* L. TRIBE, *supra* note 77, § 13-20, at 1107.

[116] 415 U.S. at 787.

[117] 440 U.S. 173 (1979).

[118] 440 U.S. at 186. *See* L. TRIBE, *supra* note 77, at 1107.

[119] 440 U.S. at 186. *See also* L. TRIBE, *supra* note 77, at 1108.

[120] 460 U.S. 780 (1983).

election.[121] Although *Anderson* was a victory on the merits for third parties, the Court abandoned the strict scrutiny approach entirely. Instead, the Court adopted an "open-ended balancing approach"[122] which called for weighing "the character and magnitude of the asserted injury" against "the precise interests put forward by the State as justifications."[123]

In finding for the plaintiff, the Court took a realistic approach to the problem of ballot access, noting that campaigns are not "static" and that changing issues will create opportunities for new candidates well after Ohio's early filing deadline. The Court also recognized that the early deadline hindered efforts to gather signatures, as volunteers are difficult to recruit, publicity and contributions hard to obtain, and voter apathy difficult to overcome so far in advance of the election.[124] Voter rights were impinged upon because "a late-emerging presidential candidate outside the major parties, whose position on the issues could command widespread community support, is excluded . . . ."[125] Although speaking only to early deadlines, and not to a state's interest in a preliminary showing of support, *Anderson*'s recognition of the practical difficulties in petitioning offered renewed hope to proponents of open ballot access.

In the next ballot-access case to reach the Court, *Munro v. Socialist Workers Party*,[126] the Court employed the less rigorous balancing test of *Anderson* but did not demonstrate the same realistic approach to reviewing the obstacles facing third-party and independent candidates. In *Munro*, the Court upheld a Washington statute that required a minor party seeking a place on the general election ballot to nominate a single candidate prior to the state primary election. The statute then required the third-party candidate to receive at least one percent of the votes cast for that office in the primary in order to appear on

---

[121] *Id.* at 782–83. Petitioner John Anderson did not challenge the state's signature requirements, which by this time required individual candidates to submit just five thousand signatures. *See id.* at 783 n.1.

[122] L. TRIBE, *supra* note 77, § 13-20, at 1108.

[123] 460 U.S. at 789. The Court based its argument on the first and fourteenth amendments generally, but it did not explicitly engage in a separate equal protection analysis. The Court did state, however, that it "rel[ied] . . . on the analysis in a number of our prior election cases resting on the Equal Protection Clause of the Fourteenth Amendment." *Id.* at 786 n.7.

[124] *Id.* at 790.

[125] *Id.* at 792. *See also* L. TRIBE, *supra* note 77, § 13-20, at 1109.

[126] 479 U.S. 189 (1986).

the general election ballot.[127] Central to the Court's holding was an equation of the primary ballot with the general election ballot. Finding that the Washington statute "virtually guarantees . . . candidate access to a statewide ballot,"[128] the Court found no burden on first or fourteenth amendment rights.

The most devastating part of the *Munro* opinion for third parties was the Court's refusal to consider empirical evidence in determining whether the state's ballot-access restrictions did, in fact, address a legitimate state interest. The U.S. Court of Appeals for the Ninth Circuit had invalidated the Washington law in part because the state had failed to show any evidence of voter confusion, ballot overcrowding, or frivolous candidacies before enacting its ballot-access scheme.[129] Between 1907 and the law's enactment in 1977, no more than six minor-party candidates had appeared at one time on a ballot for statewide office, nor more than four for any office other than governor.[130] Between the time the law was passed and the Supreme Court's decision, only one third-party candidate for statewide office had qualified for the general-election ballot.[131] The Court, however, refused to second-guess the legislature, holding that "a particularized showing" of adverse effects was not required to support a state's interest in ballot-access restrictions.[132]

Justice Marshall, joined by Justice Brennan, wrote a scathing dissent. Marshall first excoriated the Court for failing to state the level of scrutiny applied and for refusing to apply the traditional strict scrutiny test of "least drastic means."[133] Marshall next criticized the Court for equating the primary ballot with the general-election ballot, pointing out that the Court's earlier holdings had correctly characterized the primary as a "forum for continuing intraparty feuds" and the general election as the "arena where issues are sharpened, policies are hotly debated,

---

[127] *Id.* at 191. Persons voting had to choose between voting in the major party primary or voting for the third party.

[128] *Id.* at 199.

[129] Socialist Workers Party v. Secretary of State, 765 F.2d 1417, 1420 (9th Cir. 1985).

[130] 479 U.S. at 203.

[131] *Id.* at 206 (Marshall, J., dissenting). That candidate was the 1984 Libertarian Party candidate for state Treasurer. Both the Republican and Democratic primaries were uncontested. *Id.* n.2. The majority, however, argued that 36 of 40 minor-party candidates for non-statewide office and four of five independents for statewide office had qualified for the general election ballot under the statute. *Id.* at 197 n.11.

[132] *Id.* at 194–95.

[133] *Id.* at 203 (Marshall, J., dissenting).

and the candidates' positions are clarified."[134] Finally, the dissent criticized the Court for ignoring the historical record. Earlier cases, including *Jenness* and *American Party*, had upheld statutes in part because the states were able to show that minor-party candidates had qualified in the past. "Under this reasoning," Marshall wrote, "the validity of ballot access limitations is a function of empirical evidence."[135] The central idea was that the definition of a "reasonable" restriction can only be determined in light of empirical evidence, which the Court had announced it would no longer consider.

The *Munro* holding that no particularized showing of state interest is required, combined with the *Jenness* test that virtually any past success by third parties in obtaining ballot status establishes the legitimacy of a ballot-access statute, undermines most constitutional challenges to ballot-access restrictions. In *Munro*, the Court recognized compelling state interests in imposing restrictions, which future plaintiffs are unable to challenge empirically. At the same time, plaintiffs will rarely be able to show an impermissible burden, for so long as even one third-party candidate has previously met a state's requirement, the law would meet the *Jenness* test. If the Court adheres to these rulings, ballot-access laws will be unassailable.

Ironically, while the Court has moved away from strict scrutiny and toward greater deference to state legislatures in reviewing signature requirements, it has consistently struck down other barriers to access. For example, although many states still have laws on the books requiring loyalty oaths, these have not been enforced since the Court found them unconstitutional in 1974.[136] Similarly, the Court has ruled that states must provide an indigent candidate with an alternative to, or a waiver of, filing fees.[137] Finally, in *Moore v. Ogilvie*, the Court held that laws requiring signatures to be distributed evenly across counties violated the principle of one man, one vote.[138]

---

[134] *Id.* at 201 (citing *Storer v. Brown*, 415 U.S. at 735). Thirty-two major party candidates—18 Democrats and 14 Republicans—appeared on the primary ballot, a fact which Marshall used to question the seriousness of the state's efforts to avoid voter confusion. 479 U.S. at 203–04 (Marshall, J., dissenting). *See also* L. TRIBE, *supra* note 77, § 13-21, at 1110 n.65 ("[T]he Court's treatment of primary elections seems misguided.").

[135] 479 U.S. at 205.

[136] Communist Party v. Whitcomb, 414 U.S. 441 (1974).

[137] Lubin v. Panish, 415 U.S. 709 (1974); Bullock v. Carter, 405 U.S. 134 (1972). These cases are discussed *infra* text accompanying notes 161–170.

[138] 394 U.S. 814 (1969).

Such decisions have resulted in a body of law striking down minor burdens on third parties but leaving intact the greatest barriers to ballot access. For example, county distribution requirements are per se barred, with no "totality" analysis whatsoever, regardless of whether candidates have in the past been able to satisfy them. Nevertheless, one could certainly design a law requiring a cross-county distribution of signatures that would be far less exclusive than the provisions upheld in *Jenness*. The Court's insistence on analyzing the burden of signature barriers in terms of whether a third party can ever gain formal access to the ballot has led it to strike down relatively minor nuisances such as distribution requirements and filing fees, while upholding far more onerous burdens on the rights at stake. A state may effectively burden these rights, so long as it does so in an approved manner.

III. The Supreme Court Revisited: Coming to Grips with the Third-Party Challenge

Courts have struck down relatively minor burdens on voting and associational rights, while frequently upholding more serious burdens, because courts have operated in a world of political theory that simply does not reflect reality. This blindness is caused in large part by the reluctance of courts to examine the actual impact of ballot-access laws on voters, parties, and state interests. A re-examination of the cases in light of the realities of the American electoral process will prove that the interests of third parties and independent candidates in ballot access are more deserving of constitutional protection than the Supreme Court has yet recognized. This review will also demonstrate that the asserted state interests are less compelling than the Court has indicated.

A. *Constitutionally Protected Interests*

1. The Right to Vote Effectively

Since *Williams v. Rhodes* the courts have recognized two fundamental rights impaired by ballot-access restrictions: the right to vote effectively and the right of association.[139] The Supreme Court has never clarified precisely what is meant by the right to vote effectively, except to suggest that the right is impermissibly burdened if voters are presented with a narrow range of candidates while other parties are denied access to the ballot.[140]

At a minimum, the right to vote effectively would seem to preclude any statutory scheme which either prohibits write-in votes, as was the case in *Williams*, or fails to tabulate write-in votes.[141] Write-in votes are not usually perceived to be an "effective" means of voting.[142] Nevertheless, the right to vote ef-

---

[139] *See, e.g., Williams*, 393 U.S. at 30; *Munro*, 479 U.S. at 193; *Anderson*, 460 U.S. at 783; *Illinois State Board*, 440 U.S. at 184; *American Party*, 415 U.S. at 780; *Storer*, 415 U.S. at 729; *Jenness*, 403 U.S. at 431; Erum v. Cayetano, 881 F.2d 689, 691 (9th Cir. 1989); Manifold v. Blunt, 863 F.2d 1368, 1372 (8th Cir. 1988); Rainbow Coalition v. Oklahoma State Election Bd., 844 F.2d 740, 743 (10th Cir. 1988). *See also* Note, *Access to the General Election Ballot for Political Parties and Independent Candidates*, 88 HARV. L. REV. 1121, 1134–35 (1975) [hereinafter Note, *Access to the General Election Ballot*].

A third right—the right to candidacy—has not been recognized as "fundamental" by the Court. *See* Clements v. Fashing, 457 U.S. 957, 963 (1982) (upholding a Texas statute which restricts the rights of candidates to run for one office while holding another). *But see id.* at 977 n.2 (Brennan, J., dissenting) ("Although we have never defined candidacy as a fundamental right, we have clearly recognized that restrictions on candidacy impinge on First Amendment rights of candidates and voters."). A distinction exists between the right of states to restrict candidate eligibility for reasons unrelated to a candidate's views, party affiliation, or membership in a suspect class, and restrictions requiring a show of support based on political affiliation. The former restricts the rights of the candidate, but has relatively little effect on voters, who can presumably find a similar candidate to place on the ballot. The latter, however, deprives groups of voters of the opportunity to nominate anyone for office.

In the wake of *Williams*, some commentators suggested that the Court had recognized a right to candidacy. *See, e.g.*, Comment, *Durational Residence Requirements for Candidates*, 40 U. CHI. L. REV. 357, 369 (1973); Note, *Durational Residence Requirements for State and Local Office: A Violation of Equal Protection?*, 45 S. CAL. L. REV. 996, 1009 (1972). This position is not tenable after *Clements v. Fashing*.

[140] *Williams*, 393 U.S. at 31. *See* Note, *Access to the General Election Ballot*, *supra* note 139, at 1134–35.

[141] Many states do not routinely tabulate write-in votes. *See, e.g.*, MO. ANN. STAT. § 115.453(4) (Vernon Supp. 1990).

[142] *See Anderson*, 460 U.S. at 799 n.26; *Lubin*, 415 U.S. at 719 n.5; *Williams*, 393 U.S. at 37 (Douglas, J., concurring). In 1936, Union Party voters wishing to vote the party slate in New York would have had to write in over 47 names in the space of three minutes. Party leaders urged supporters to vote Republican rather than write in the Party name. Prohibition Party leaders urged voters not to vote at all when the Party was not on the ballot. *See* Note, *Limitations on Access to the General Election Ballot*, 37 COLUM. L. REV. 86, 98 n.84 (1937). There is also a high probability of write-in votes being declared void for error. *Id.*

Write-in votes, however, are not always in vain. In 1954 Strom Thurmond was elected to the United States Senate after a write-in campaign in South Carolina. Jackie Stump, a miners' union official running a write-in campaign, defeated a 20-year Democratic incumbent to win a seat in the Virginia legislature in 1989. The district was the site of a bitter strike by miners. Ballot Access News, Dec. 24, 1989, at 2, col. 2.

fectively may be served by write-in votes as long as the voter is allotted sufficient time, a disproportionate percentage of write-in votes are not disqualified for technical flaws, and the votes are included in official tabulations.[143] A write-in vote in such circumstances quite clearly states the voter's preference, and carries the same weight as any other vote in determining the winner. Accordingly, the right to vote effectively is reasonably well protected by the write-in system.

However, a write-in vote does not protect the right to associate for political ends. A third-party write-in candidate simply does not compete on the same footing as candidates who appear on the ballot. The Court has failed to recognize this disadvantage. We turn now to a discussion of the underprotected associational rights of third-party candidates and the voters who support them.

2. The Right of Association

Ballot-access restrictions force third-party and independent candidates to compete on consistently unequal terms with the Republican and Democratic parties. Furthermore, state petition requirements in and of themselves burden the right to associate privately, and the right to a secret ballot, whether or not the candidate ultimately appears on the ballot.

The Supreme Court has stated that the right of association is closely related to the right of free speech:

> Effective advocacy of both public and private points of view, particularly controversial ones, is undeniably enhanced by group association, as this Court has more than once recognized by remarking on the close nexus between the freedoms of speech and assembly.[144]

---

[143] There is a justifiable reluctance to put too much emphasis on the ineffectiveness of write-in votes. Without write-in voting, the right to vote effectively might come to require that any candidate supported by even one voter be granted access to the ballot. *Cf.* L. TRIBE, *supra* note 77, § 13-20, at 1103 n.12. (noting that the right of voters to vote for *particular* candidates may entail access for any candidate with the support of one voter). Such a consequence would be perceived by some observers as leading to chaos. *But see infra* notes 212–214 and accompanying text, discussing experience with independent candidates in Oklahoma, where no showing of support is required before granting ballot status. *See also* J. WIGMORE, *supra* note 26, at 53 (discussing the successful use of the Australian ballot in South Australia in the 19th century, with just two signatures required to gain a place on the ballot).

[144] NAACP v. Alabama *ex. rel.* Patterson, 357 U.S. 449, 460 (1958), *quoted in* Note, *Access to the General Election Ballot*, *supra* note 139, at 1136.

While ballot-access laws do not directly limit the right of association, as Justice Harlan wrote in his concurring opinion in *Williams*, "[b]y denying [third parties] any opportunity to participate in the procedure by which the President is selected, the State has eliminated the basic incentive that all political parties have for conducting such activities, thereby depriving [them] of much of the substance, if not the form, of their protected rights."[145] Combined with the Court's hints that write-in ballots are not an adequate substitute for ballot space, and the idea that voters are burdened if a party cannot place its candidates on the ballot, this conclusion implies a broad substantive right to associate which may not be infringed upon by the state.

However, in defining the burden ballot-access laws place on associational rights, the Court has focused narrowly on the ability of third parties to qualify for the ballot.[146] Writing for the Court in *Williams*, Justice Black defined the burden of ballot-access laws solely in terms of whether a party could be kept off the ballot entirely: "The right to form a party for the advancement of political goals means little if a party can be kept off the election ballot and thus denied an equal opportunity to win votes."[147] Justice Harlan's concurrence[148] also emphasized the inability of third-party candidates to participate in the election. Indeed, the Court has repeatedly held that if even the occasional third-party candidate can qualify for the ballot, the first amendment association rights of third-party supporters are not impermissibly burdened.[149] Further, in *Munro* the Court specifically found that no burden on associational rights existed because

---

[145] *Williams*, 393 U.S. at 41 (Harlan, J., concurring). *See also* Lawson, *How State Laws Undermine Parties*, in ELECTIONS AMERICAN STYLE, *supra* note 26, at 244 ("A party that is not able to compete in free elections cannot properly be said to be a party.").

[146] Most challenges to ballot-access restrictions, including all those to reach the Supreme Court, have been brought only after a third party has been denied access to the ballot. However, there are cases which discuss other kinds of burdens on associational rights. *See* Fulani v. State Election Board, No. 88-3122 (D. Ind. 1989) (challenging action of state election officials in placing 1988 Republican and Democratic presidential candidates on ballot although both parties filed their lists of electors after the statutory deadline); Dart v. Brown, 717 F.2d 1491 (5th Cir. 1983), *cert. denied*, 469 U.S. 825 (1984) (unsuccessful challenge to Louisiana statute placing party affiliation after the names of Democratic and Republican candidates, but not third-party candidates, in state's unitary run-off elections); Board of Election Comm'rs v. Libertarian Party, 591 F.2d 22 (7th Cir. 1979), *cert. denied*, 442 U.S. 918 (1979) (challenging rule automatically granting top two ballot positions to Democrats and Republicans).

[147] 393 U.S. at 31.

[148] *Id.* at 41 (Harlan, J., concurring).

[149] *See Munro*, 479 U.S. at 196–97; *American Party*, 415 U.S. at 783–84; *Jenness*, 403 U.S. at 439.

third parties were given easy access to the primary ballot, though not the general-election ballot.[150]

The Court's narrow reading of the right of association has failed to address even minimally the real equal protection and first amendment claims of minor-party and independent candidates. Signature requirements burden associational rights in other ways short of complete deprivation of a spot on the ballot. The cost and effort involved in meeting petition requirements are such that a candidate's successful hurdle of ballot-access restrictions often amounts to little more than a Pyrrhic victory. One observer, after researching twentieth century third-party presidential campaigns, concluded, "almost every candidate who had run as a third-party or independent candidate for the Presidency had basically gotten submerged in the problems of getting onto the ballot, and exhausted his or her resources by September or October, with nothing left to run a real campaign."[151] The drain extends beyond cash resources to include volunteer time and enthusiasm lost in the thankless job of petitioning. Parties are also harmed by the fact that the lack of guaranteed ballot status makes it more difficult to attract volunteers, and in most states, impossible to register voters under the party name.[152]

The Court's ballot-access jurisprudence has overlooked a second aspect of associational rights: the right to associate privately. In *NAACP v. Alabama ex rel. Patterson*,[153] the Supreme Court recognized that the right to associate can sometimes be greatly harmed by eroding the confidentiality of the association. This right of privacy in association was extended to political organizations in *Brown v. Socialist Workers '74 Campaign Committee*.[154] In *Brown*, the Court ruled that Ohio could not require the Socialist Workers Party to report the names and addresses of campaign contributors or recipients of campaign disbursements, noting that the District Court had found "substantial evidence of both governmental and private hostility toward and harrassment of SWP members and supporters."[155] The Court held that forced disclosure would unduly burden the first amend-

---

150 479 U.S. at 199.

151 Frampton, *Challenging Restrictive Ballot Access Laws on Behalf of the Independent Candidate*, 10 REV. OF L. & SOC. CHANGE 131, 134 (1981).

152 Hocker, *supra* note 5, at 127.

153 357 U.S. 449 (1958).

154 459 U.S. 87 (1982).

155 *Id.* at 98–99.

ment rights of members of the organization. Ballot-access requirements that force supporters to place publicly their names and addresses on a candidate's or party's petition naturally raise the same constitutional issues. *Brown* is not dispositive of those issues, even with respect to the Socialist Workers Party, since the state's interest in limiting the number of candidates on the ballot may be greater than the state interest in disclosure of campaign financing, thus justifying the burdens imposed.

Such government and private harassment is not a phenomenon limited to the most "radical," smaller third parties, such as the Socialist Workers Party. In 1948, for example, newspapers in New Haven, Pittsburgh, Boston, Milwaukee, and Cleveland discouraged support for Henry Wallace—a former Cabinet member and Vice President of the United States—by publishing the names, addresses, and occupations of people who signed his ballot petitions.[156] During the presidency of Woodrow Wilson, Socialist Party members were tarred and feathered and their meetings broken up by vigilantes, despite the fact that the party held majorities in many city councils and was represented in the U.S. Congress.[157] Although such egregious intimidation is uncommon today, it still is "an extraordinary act for Americans to vote for a third party."[158] In order to vote for the candidate of their choice "they must often endure ridicule and harassment from neighbors and friends . . . ."[159]

Finally, requiring a candidate to show significant support through a petition drive also burdens the constitutional right to a secret ballot. Though never explicitly acknowledged by the Supreme Court, this right has been recognized by lower courts,[160] and was a primary motivation behind the adoption of the Australian ballot. Even in states which do not require petition signers to state that they are members of the party whose petition they sign, having to sign a public petition tarnishes the signer's right to a secret ballot. This would not be true if one assumed that signing a petition was not a show of support for the party, an assumption which, if accepted, would cast doubt

---

[156] K. Schmidt, Henry A. Wallace, the Quixotic Crusade 133–34 (1948).

[157] J. Weinstein, The Decline of Socialism in America, 1912–1925, at 140–45 (1967).

[158] S. Rosenstone, *supra* note 9, at 3.

[159] *Id.*

[160] *See, e.g.*, Anderson v. Mills, 664 F.2d 600, 608 (6th Cir. 1981); Buckley v. Valeo, 519 F.2d 821, 867 n.117 (D.C. Cir. 1975), *rev'd on other grounds*, 424 U.S. 1 (1976); Libertarian Party v. Beermann, 598 F. Supp. 57, 60 (D. Neb. 1984).

on whether the state had any interest in requiring signatures as a preliminary showing of support.

3. The Filing Fee Cases: An Anomaly

To determine that a third party and its supporters have not been burdened merely because their candidates appeared on the ballot is to give a hollow meaning to both the equal protection clause and the first amendment rights that the Court claims to have recognized. Indeed, in the past the Court has given full protection to related rights based on the equal protection clause and the first amendment. In cases involving other restrictions on the electoral process, the Court concluded that the rights to associate and to vote effectively were overly burdened. A comparison of the cases involving signature requirements with two non-signature cases, *Bullock v. Carter* and *Lubin v. Panish*, demonstrates the Court's inconsistent approach in its analysis of the burdens placed on participation in the electoral process.

In both *Bullock* and *Lubin*, the Court firmly declared that a mandatory filing fee, with no provision for waiver or alternative means of ballot access, is constitutionally impermissible. In *Bullock*, the Court struck down a Texas statute which required candidates in the major-party primaries to help to finance the cost of the primary through filing fees that ran as high as $8,900.[161] The Court found that such sizeable fees could not be presumed to be satisfied from the candidate's personal resources.[162] To the extent that the system required the candidate to rely on contributions to pay the fees, the Court recognized that it burdened the franchise and fell "with unequal weight on voters, as well as candidates, according to their economic status."[163] The discriminatory nature of the filing fee requirement led the Court to scrutinize the statutes closely.[164]

The state advanced two interests to defend the statute—limiting ballot size and collecting revenue. The Court concluded that the state's interest in collecting revenue lacked a showing of necessity, particularly as it was the state, and not the candidates, which benefited from the financing of the primary through

161 405 U.S. 134 (1972).
162 *Id.* at 143.
163 *Id.* at 144.
164 *Id.*

filing fees.[165] The state's asserted interest in keeping the ballot to a manageable size also failed. The Court found the payment of a filing fee to be "extraordinarily ill-fitted to that goal."[166] Fees would not keep frivolous but wealthy candidates off the ballot, yet might keep serious but poorly financed candidates from being presented to the voters. The *Bullock* decision framed the issue as an infringement more on voting rights than on associational rights. This approach has legitimacy given that write-in votes were prohibited in the primaries.

In *Lubin*[167] the Court struck down a California statute mandating much more modest filing fees than those required in *Bullock*. The Court again insisted that the state show that no less restrictive alternative existed before burdening voting rights. Unlike *Bullock*, however, the Court emphasized voters' associational rights, suggesting that those rights required a system that mandated access to the ballot for all candidates.[168] Additionally, the Court indicated that any filing fee, even one as low as a single dollar, would be suspect if it did not provide an alternative means of ballot access for indigent candidates.[169]

Bullock and *Lubin* give a particularly curious twist to the Supreme Court's ballot-access jurisprudence when compared with *Williams* and *Jenness*. The Court's willingness to strike down filing fees seems entirely incongruous with its toleration of significant signature requirements as a permissible burden on first and fourteenth amendment rights. This inconsistency seems to be based on the following theory: (1) the state has a legitimate interest in requiring some showing of public support, or seriousness, before placing a candidate on any ballot; (2) a candidate with no money might have significant public support, whereas a well-financed candidate may not; and (3) filing fees absolutely bar an impoverished candidate from the ballot and constitute an impermissible burden, but signatures can be gathered at no cost, so all but the most draconian signature requirements keep the political system open to challenge.[170]

---

[165] *Id.* at 147–48. Though the Court did not use the term "least drastic alternative," this was the standard applied.

[166] *Id.* at 146.

[167] 415 U.S. 709 (1974).

[168] *Id.* at 713.

[169] *Id.* at 714.

[170] *Lubin* suggests that a petition alternative to filing fees would be acceptable. 415 U.S. at 718.

Signature requirements are not, however, financially insignificant. The cost of gathering petition signatures, even for an extremely popular third-party or independent candidate such as George Wallace in 1968 or John Anderson in 1980, is quite substantial. Over thirty years ago, the estimated cost for a third party to get on the ballot in California was $100,000.[171] Recent third party experience shows that the cost of obtaining nationwide ballot status, quite apart from the time spent by campaign volunteers, is approximately $1.5 million, or more than one dollar per signature required.[172] This cost would be significantly higher if all states had enacted laws as stringent as those upheld by the Court in *Jenness v. Fortsen*. Given these numbers, it is ludicrous to suggest that a candidate who could not pay a one dollar filing fee, or even an $8,900 filing fee, could successfully conduct a major petition drive to appear on the ballot.[173]

---

[171] Christian Nationalist Party v. Jordan, 49 Cal.2d 448, 318 P.2d 473 (1957).

[172] The New Alliance Party, which obtained ballot status in all 50 states and the District of Columbia in 1988, spent between $1.4 million and $1.5 million on ballot access, gathering a total of over 1.1 million signatures on state nominating petitions. This was approximately one half of the Party's total campaign expenditures, and does not include volunteer time. Given the *Bullock* emphasis on the burden of filing fees on poorer voters, it is interesting to note that the New Alliance Party draws its core support from among the urban poor, and relied heavily on volunteer labor in conducting its petition drives. Telephone interview with David Belmont, Ballot Access Coordinator for the New Alliance Party (Feb. 5, 1990).

Andre Marrou, Director of Project 51-'92, a Libertarian political action committee that lobbies for less restrictive ballot laws and has conducted ballot drives on behalf of the Libertarian Party, reports that his organization budgeted $12,000 to $15,000, or $1.20 to $1.35 per signature, for a petition drive in Nevada in January 1990. For a drive conducted in North Carolina in January 1990, Project 51-'92 budgeted $69,000 to $84,000, or approximately $1.30 to $1.60 per signature required. Although the Libertarian Party has automatic ballot status in many states based on past election results, Marrou estimates the cost to appear on all 51 ballots in 1992 will range from $1 million to $1.5 million. Letter from Andre Marrou to Brad Smith (Jan. 26, 1990) (on file at the HARV. J. ON LEGIS.).

Costs in a petition drive typically include paid labor, lodging and vehicles for volunteers and staff, general administration, notarization of petitions, printing, travel, and fundraising to cover the other costs of the drive. Belmont interview, *supra*.

[173] This argument was rejected in Andress v. Reed, 880 F.2d 239 (9th Cir. 1989) (filing fee of $1,702 upheld over plaintiff's claim that cost to gather 10,000 signatures—the alternative provided by statute—would be more than the filing fee).

In *Storer*, Justice White, writing for the Court, found that a requirement of 325,000 signatures gathered in 24 days was not an "impossible burden," because "1000 canvassers could perform the task if each gathered 14 signers a day." 415 U.S. at 740. Justice White repeated this type of arithmetic in *American Party*, finding that 100 canvassers gathering four signatures a day for 55 days would meet the Texas requirement. 415 U.S. at 786. Justice White's figures are not quite accurate because he did not allow for the fact that guaranteeing enough good signatures usually requires at least a twenty-five percent safety margin of "raw" signatures. Marrou, *supra* note 172. Accepting Justice White's logic, however, one might just as well conclude that the statute challenged in *Lubin* posed no barrier because 1000 panhandlers collecting just three cents per day could obtain the necessary filing fee in 24 days; *Bullock's* $8,900 fee would have required 100 panhandlers to collect $1.62 each per day over 55 days.

Even assuming that there may be some petition drives, typically for local office, which might be small enough to be accomplished by the candidate alone or with a handful of supporters, it is not true that a petition requirement more equitably fulfills the state's legitimate needs than a filing fee. Signatures are not good indicators of public support. In fact, the number of signatures gathered usually has little to do with popularity. The vast majority of signatures in a petition drive are not gathered from supporters of the candidate or party, but from the public at large, by petitioners working county fairs or patrolling the sidewalks in front of shopping malls and post offices.[174]

Given the lack of attention the media pays to third parties, and the fact that petitioning must usually be done well in advance of the election, even educated voters usually know little or nothing of the candidate whose petition they sign. Most people who sign do so out of a simple conviction that every candidate deserves a chance to be on the ballot. Thus the ability to gather signatures may bear little relation to public support. Yet it does not follow that third-party candidates with relatively few openly committed supporters are wholly without electoral appeal. For example, in January 1990 the Libertarian Party reported that its national membership had reached an all-time high of 7907,[175] yet the party's presidential candidates have received as many as 921,000 votes.[176]

Indeed, a petition drive is no more likely to correlate with a candidate's ultimate support at the ballot box than is his or her ability to pay a filing fee. A candidate who cannot raise the minimal filing fee struck down in Lubin is unlikely to be able to mount a significant campaign or to garner much popular support. Conversely, a wealthy candidate who can afford to deliver his or her message through the media is likely to develop some level of public support by election day, even if he or she has little or no support at the time of the filing deadline.[177]

---

[174] By a 5-4 vote, the Supreme Court recently held constitutional a Post Office regulation banning first amendment activities on Post Office sidewalks. U.S. v. Kokinda, 110 S. Ct. 3115 (1990). The fact that the nation's two largest third parties, the Libertarian Party and New Alliance Party, filed amicus briefs urging the Court to strike down the ban indicates the importance of petitioning in such locales to third parties and independent candidates.

[175] *Membership at All-Time High*, Libertarian Party News, Jan. 1990, at 4, col. 3.

[176] Libertarian presidential vote totals have been: 1988, 432,179; 1984, 228,314; 1980, 921,299; 1976, 173,011. R. SCAMMON & A. MCGILLIVRAY, 18 AMERICA VOTES 6, 8, 10, 12 (1989).

[177] The high cost to third parties of meeting ballot-access signature requirements also

The net result of the Court's holdings, then, has been to strike down those state restrictions least likely to burden the rights of voters and parties, while upholding restrictions that pose a far greater barrier to ballot access. The Court has recognized the right to associate and has implied that it is a broad right to associate effectively. But it has failed to provide constitutional protection to those interests which undergird associational rights in the political arena: the ability to contest elections on equal footing, the right to privacy in association, and the right to a secret ballot.

B. *Compelling State Interests*

In support of ballot-access restrictions, states have asserted compelling interests in assuring majority winners, maintaining political stability, avoiding voter confusion and apathy, discouraging frivolous candidates and avoiding clogged voting machinery, and generally assuring fair and honest elections.[178] With the possible exception of the general obligation to assure fair, honest elections, each of these objectives necessarily requires that the number of candidates on the ballot stay within some reasonable number. However, significant practical barriers to the proliferation of candidacies and parties exist independent of ballot-access restrictions. A review of these barriers is necessary before discussing the asserted state interests.

Without doubt, the greatest formal obstacle to the emergence of third parties is the U.S. system of single-member plurality districts.[179] Because a party gets nothing until it is able to carry at least a plurality in the district covered by the election, the system imposes severe penalties on a party hoping to grow over

---

emphasizes the hollowness of the Court's insistence on analyzing protected interests purely in terms of qualifying for the ballot. Statutorily imposed expenditures of $1.5 million, *see* Belmont, *supra* note 172, when applied only to certain candidates and groups of voters on the basis of their perceived popularity at a point well before election day, certainly raise equal protection, if not first amendment, problems long before those candidates are formally denied access to the ballot.

[178] Note, *Fairness in the Election Arena: Congressional Regulation of Federal Ballot Access*, 32 N.Y.L. SCH. L. REV. 909–10 (1987) [hereinafter Note, *Fairness in the Election Arena*] (citing *Williams*, 393 U.S. 23; *Anderson*, 460 U.S. 780; *American Party*, 415 U.S. 767; *Storer*, 415 U.S. 724).

[179] S. ROSENSTONE, *supra* note 9, at 16–18; Lawson, *supra* note 145, at 243–44; D. RAE, POLITICAL CONSEQUENCES OF ELECTORAL LAWS 92 (1971).

time. The system also creates a powerful incentive for parties to coalesce until just two candidates are left in the running.[180]

A second formal barrier is the Federal Election Campaign Act of 1974 ("FECA"),[181] which limits the amount of money any single individual can give to a candidate,[182] requires disclosure of campaign contributors,[183] and provides federal funds to Republican and Democratic presidential candidates. FECA has been described as "a major party protection act."[184] In addition to pumping millions into the Democratic and Republican parties, FECA contains disclosure provisions that have a chilling effect on third-party fund raising, and its $1,000 limit on campaign contributions makes it impossible for a "sugar daddy" to provide seed money for a fledging party. In sum, FECA "ensures a large gap between the financial resources available to major and minor parties."[185]

In addition to these formal barriers, third parties in America face a variety of informal obstacles limiting their desirability as electoral vehicles. These include chronic money shortages, poor press coverage, lack of patronage to offer supporters, and voter attachment to the two-party system.

Even before FECA, third-party candidates suffered enormous financial disadvantages. The best funded third-party candidate for president was former President Theodore Roosevelt in 1912. He spent just sixty percent of the average amount spent by the Republican and Democratic candidates. The typical third-party candidate is outspent by ratios of fifty to one or more.[186]

---

[180] In presidential elections, this effect is amplified by the electoral college, for a party must not only capture a winner-take-all state, but it must also capture enough states to win with an absolute majority. Only by concentrating strength in a particular region have third parties been able to survive in countries with single-member districts. D. RAE, *supra* note 179, at 94–95.

[181] 2 U.S.C. § 431–456 (1988).

[182] *Id.* § 441a.

[183] *Id.* § 434.

[184] S. ROSENSTONE, *supra* note 9, at 26.

[185] *Id.* at 27. *See generally id.* at 25–27. Although the act was upheld in Buckley v. Valeo, 424 U.S. 1 (1976), there is still room for the Court to reverse its earlier holding. In *Buckley*, the Court rejected speculative arguments that the act worked invidious discrimination against independent and third-party candidates, *id.* at 102, but held open the possibility that such discrimination might be shown in the future. *Id.* at 97 n.131. The difficulties John Anderson had in obtaining loans against his expected post-election matching funds might suggest that the challenge could be raised again. Rada, Cardwell & Friedman, *Access to the Ballot*, 13 URB. LAW. 793, 808 (1981).

[186] S. ROSENSTONE, *supra* note 9, at 27–29. Had FECA been in effect, Roosevelt would probably have been worse off, for he relied heavily on a handful of contributors whose contributions far exceeded FECA's $1,000 maximum. *Id.* at 27–29 n.10.

Third-party candidates also receive far less free media coverage than their major-party rivals. They are routinely excluded from televised debates.[187] In 1988, the News Election Service, a major source of election results for all three television networks and several major newspapers, adopted a policy of not reporting third-party vote totals.[188] When third parties do receive media coverage, it is often hostile. Even when it is not, the tendency of the press to focus on "horse-race" aspects of the campaign means that coverage of third-party candidates is heavily focused on the the ultimate futility of their candidacy.[189] Indeed, "[w]hen voters support third-party candidates, they do so in spite of, not because of, the media's coverage of their campaigns."[190]

The strongest of the informal barriers to proliferating candidacies is voter attachment to the two-party system in general,[191] and to the Democratic and Republican parties in particular. Even in today's era of weakened party allegiance, a solid majority of Americans feel deep familial, social, and political ties to one of the two major parties.[192] In fact, a 1977 national survey of eighth-grade students found that half thought it was illegal to organize a new political party.[193]

All of these factors, both formal and informal, are mutually reinforcing. Public attachment to the established parties makes it difficult to raise funds; lack of funds limits a third-party campaign's reach, making winning less likely; this results in reduced

---

[187] *See* Fulani v. League of Women Voters Educ. Fund, 882 F.2d 621 (2d Cir. 1989) (New Alliance Party successfully showed harm from exclusion from debates); Fulani v. Brady, 729 F. Supp. 158 (D.D.C. 1990) (New Alliance Party challenged tax-exempt status of Commission on Presidential Debates on grounds that Commission had violated its requirement to remain non-partisan but Party was found to lack standing).

[188] Ballot Access News, Nov. 27, 1989, at 1, col. 1. According to Ballot Access News, this policy caused the News Election Service ("NES") some embarrassment in the 1989 elections in New York City. In a three-way race for city council, New Alliance candidate Pedro Espada finished second with 42.5% of the vote. NES reported the vote totals for the first and third place candidates, but not Espada. Several New York newspapers which rely on NES for election results declined to report on Espada's vote totals even after being contacted by the New Alliance Party. *Id.* Responding to an inquiry from U.S. Rep. Edward Markey (D-Mass.), NES stated that "this failure . . . was a mistake." *Id.* Feb. 12, 1990, at 3, col. 3.

[189] S. ROSENSTONE, *supra* note 9, at 33–37.

[190] *Id.* at 37.

[191] V.O. KEY, *supra* note 6, at 209–10.

[192] *Id. See also* Note, *Legal Obstacles to Minority Party Success*, 57 YALE L.J. 1276, 1287 n.43 (1948).

[193] 41 *The Progressive* 14 (Mar. 1977). One can speculate as to whether this is cause or effect, and whether this represents a failure of civic education or a powerful, yet subtle, success.

press coverage and further pressure on voters to gravitate to their preferred major party candidate in winner-take-all districts; and so on. The important point for our purposes is that these various factors pose a formidable barrier to third-party success, making it unlikely that third parties will threaten political stability or result in large numbers of bare plurality winners. Furthermore, because these factors are known to would-be independent and third-party candidates and supporters in advance, they are a significant deterrent to frivolous candidacies and a powerful incentive to serious candidates and voters to go outside the two major parties only as a last resort.[194]

Given the enormous formal and informal constraints on third-party candidacies, the interests states assert to justify additional barriers to third-party ballot status do not warrant the burdens placed on constitutional rights. States can achieve their ends more effectively with less restrictive means, especially since ballot-access laws often do not achieve the state's intended results.

The state's asserted interest in assuring majority winners was quickly dismissed by the *Williams* Court[195] and is surely not sufficiently compelling to justify burdens on first and fourteenth amendment rights. The underlying objective of this interest appears to be to assure public mandates for and confidence in elected officials. However, public legitimacy is based on process—and the belief that the process is fair—as much as vote totals. Thirteen U.S. Presidents have been elected by plurality votes in fifteen elections, with no apparent loss in legitimacy.[196] Low voter turnout[197] means that a candidate receiving a majority of the votes cast usually will not have been elected by a majority of the total eligible voters. Furthermore, the third-party candidates most likely to prevent a majority winner from emerging are those who will gather a significant percentage of the vote. Yet candidates with significant support are precisely those whom it would be difficult to keep off the ballot constitutionally,

---

[194] *See* S. ROSENSTONE, *supra* note 9, at 215–16 ("Third parties are a response to major party failure."); D. MAZMANIAN, *supra* note 8, at 27 ("The leading precondition for a significant third party vote is severe political crisis.").

[195] *See* 393 U.S. at 32. *Williams* is discussed *supra* text accompanying notes 54–73.

[196] U.S. ELECTIONS, *supra* note 4, at 321. Grover Cleveland and Woodrow Wilson were each elected to two terms by pluralities.

[197] Voter turnout in the United States is typically in the range of 60% during presidential election years, and lower in other years. Burnham, *The Turnout Problem*, in ELECTIONS AMERICAN STYLE, *supra* note 26, at 97, 113–14.

even under the present system.[198] In any event, the states can assure majority winners without ballot-access restrictions. Runoff elections and non-partisan primaries are two methods which allow third parties to test their strength, and, if warranted, to join forces with the more successful party nearest their views before the general election.

The second state objective, political stability, is sometimes described as preventing factionalism and party splintering. A two-party system does foster political stability, which is clearly a compelling interest of the state.[199] But the state may have less drastic methods available to accomplish these ends. For example, party splintering and candidacies provoked by intra-party feuds are readily deterred by "sore loser" laws, such as that challenged in *Storer v. Brown*.[200] Furthermore, like the professed interest in assuring majority winners, the interest in maintaining a two-party system through ballot-access laws fails because those parties most likely to destabilize the system are those with the greatest constitutional claim to a place on the ballot.[201] Finally, given the stability of the system in the years before the Australian ballot came into existence, the powerful barriers posed by single-member districts, and equally powerful informal barriers, it is difficult to explain why additional statutory barriers are needed.

Still more suspect is the asserted state interest in preventing voter confusion and apathy. Even assuming that less restrictive access provisions will result in large numbers of candidates,[202] the argument remains weak. Although there is some evidence to suggest that voter turnout drops during periods of significant third-party activity, most studies of voting behavior point to apathy as a cause, not an effect, of third-party proliferation.[203] While a ballot may at some point become so cluttered as to confuse or alienate voters, any statutory scheme that seems to limit the ballot routinely to as few as three or four candidates should be suspect.[204] As most voters identify with one of the

---

[198] *See* Note, *Access to the General Election Ballot*, *supra* note 139.

[199] *See* Jardine, *supra* note 112, at 304.

[200] 415 U.S. 724 (1974). *See supra* text accompanying notes 106–114.

[201] *See Williams*, 393 U.S. 23 (1968) (striking down an Ohio statute which kept George Wallace off the ballot despite his significant popular support), discussed *supra* text accompanying notes 54–73.

[202] *See infra* text accompanying notes 208–214.

[203] *See, e.g.*, D. MAZMANIAN, *supra* note 8, at 77–81.

[204] Jardine, *supra* note 112, at 305.

major parties, a ballot listing several minor parties will not normally be confusing.[205] Voters can usually select their candidate from a list of a dozen as easily as from a list of three, because most people make their choices long before entering the polls.[206] Further, excluding groups from the ballot may create a dangerous motivation for those groups to attempt to take over one of the major parties. Such a result would be at least as likely to cause voter confusion and political instability as the presence of identified third parties on the ballot. This problem attracted significant attention in 1986 when followers of Lyndon LaRouche won the Illinois Democratic primaries for Secretary of State and Lieutenant Governor. As a result, the Democratic gubernatorial nominee, Adlai Stevenson, whose name is virtually synonomous with the Democratic Party in Illinois, organized the Illinois Solidarity Party to conduct his campaign.[207]

One of the more convincing state interests, at least abstractly, is the need to avoid overwhelming voting machinery with large numbers of frivolous or insignificant candidates.[208] Particularly in jurisdictions using voting machines, an extremely large number of candidates has the potential to create severe administra-

---

[205] Note, *Access to the General Election Ballot*, *supra* note 139, at 1137–38.

[206] H. PENNIMAN, SAIT'S AMERICAN PARTIES AND ELECTIONS 367 (5th ed. 1952). *But see generally* Barton, *The General-Election Ballot: More Nominees or More Representative Nominees?*, 22 STAN. L. REV. 165 (1970).

[207] N.Y. Times, July 27, 1986, at A12, col. 1. LaRouche had first tried to enter politics through the U.S. Labor Party, and began to infiltrate the Democrats only after his third party was unable to appear on most state ballots. For an overview of LaRouche's early career, see D. KING, LYNDON LAROUCHE AND THE NEW AMERICAN FASCISM (1989). LaRouche supporters continue to trouble the Democratic Party. A LaRouche supporter captured the Democratic nomination for Congress in Michigan's Tenth District in 1990. It is generally a Republican district, although a Democrat won the Congressional seat as recently as 1982. Ballot Access News, Oct. 9, 1990, at 4, col. 2.

Another politician who made his first electoral efforts through a third party is David Duke, the Klansman who came to national attention when he was elected to the Louisiana state legislature as a Republican in 1989. *See* N.Y. Times, Feb. 26, 1989, § 4 (Week in Review), at 7, col. 1. In 1990 Duke ran for the U.S. Senate as a Republican. Fear of a Duke victory led the other Republican candidate in Louisiana's single-ballot Senate primary to withdraw from the race and transfer his support to the Democratic incumbent, J. Bennett Johnston. *Id.*, Oct. 5, 1990, at A1, col. 5. Although Johnston won re-election, Duke gathered 44% of the vote. *Id.*, Nov. 8, 1990, at B8, col. 3. Duke had been the Populist Party candidate for President in 1988. *Id.*, Nov. 22, 1988, at B6, col. 6.

In New York, New Alliance Party activists Pedro Espada and Sandra Love were elected Democratic Party district representatives in 1990. In Maryland four Libertarian Party members won Republican primaries for the state legislature in 1990. After the primary they began publicizing their Libertarian affiliations. No third party has appeared on the Maryland ballot since the current ballot-access statute was passed in 1971. Ballot Access News, Oct. 9, 1990, at 4, col. 2.

[208] *See* Jardine, *supra* note 112, at 306.

tive problems for the state.[209] On the other hand, in *Bullock v. Carter* the Court, while recognizing a legitimate state interest in conserving tax dollars, found the interest insufficient to justify burdening the franchise.[210] Although it may in some cases add to the cost of holding elections, states can arrange for voting mechanisms sufficient to handle over two dozen candidates.[211]

But a relaxation in ballot-access requirements would not necessarily create a flood of candidates. States which have minimal ballot-access requirements have not experienced any inordinate problems with ballot clutter. Oklahoma, for example, does not require any showing of voter support before placing independent candidates for statewide office, U.S. House, and U.S. Senate on the ballot.[212] It is not generally considered a one-party state—in 1990 it was represented in Congress by four Democrats and two Republicans, with one Republican and one Democratic Senator, a Republican Governor, and a Democratic Lieutenant Governor. Yet no more than eight, and since 1934 no more than six, candidates have ever appeared on any Oklahoma general election ballot for any of these offices.[213] Since 1950, 77.2% of all such races have been contested by no more than two candidates, and 93.7% by no more than three candidates.[214]

---

[209] *See* Rockefeller v. Socialist Workers Party, 314 F. Supp. 984 (S.D.N.Y. 1970), *aff'd mem.*, 400 U.S. 806 (1970), discussed *supra* text accompanying notes 99–101.

[210] 405 U.S. 134, 144–45, 147–49 (1972).

[211] Thirty-three candidates (32 of whom were Democrats and Republicans) appeared on the Washington state primary ballot that provided the backdrop to the *Munro* decision. 479 U.S. at 203–04.

[212] Independent candidates are given the option of filing a petition signed by five percent of the registered voters in the electoral district, or paying relatively minor filing fees ranging from $200 for most offices to $1,500 for governor. OKLA. STAT. ANN. tit. 26, § 5-112 (West Supp. 1990). Filing fees at this level are more of a nuisance than a substantive barrier to a serious candidate. Prior to 1967, the law was even more lenient, allowing independents to qualify merely by submitting a petition signed by two registered voters. *See* OKLA. STAT. ANN. tit. 26, § 5-112 (West 1951 and Supp. 1968). The waiver of the five percent petition requirement is not available to third-party candidates, OKLA. STAT. ANN. tit. 26, § 1-108 (West Supp. 1990), but in such circumstances third-party candidates often run as independents. For a discussion of Oklahoma's law governing third-party ballot access, see *infra* text accompanying notes 229–236.

Other states with particularly lenient ballot-access requirements include Mississippi, which merely requires a third party to be organized, MISS. CODE ANN. § 23-15-1061 (Supp. 1989), and Tennessee, which places independent candidates for statewide office on the ballot with the submission of just 25 signatures. TENN. CODE ANN. § 2-5-101(b)(1) (Supp. 1989).

[213] *See* OKLAHOMA STATE ELECTION BOARD, 1 OKLAHOMA ELECTIONS: STATEHOOD TO PRESENT (1988) [hereinafter OKLAHOMA ELECTIONS]; OKLAHOMA STATE ELECTION BOARD, ELECTION RESULTS AND STATISTICS 1988 (1988) [hereinafter 1988 OKLAHOMA RESULTS].

[214] *See* OKLAHOMA ELECTIONS, *supra* note 213, at D-363 to D-611; 1988 OKLAHOMA RESULTS, *supra* note 213, at 61–65.

This finding has implications that go beyond the question of clogging election machinery, for if the absence of meaningful ballot-access restrictions does not result in an explosion of candidacies, then the state interest in majority winners, political stability, and preventing voter confusion are also significantly reduced. At a bare minimum, the evidence indicates that most states could accomodate any legitimate administrative concerns with far less restrictive laws than they now employ.

Finally, there is the interest, never fully defined by the Court, in assuring fair and honest elections. While this certainly sounds like a legitimate state interest, it too turns out to be unsubstantiated and spurious. In *Manifold v. Blunt*,[215] for example, a federal court upheld a Missouri statute requiring third parties to submit lists of presidential electors earlier than the Republicans and Democrats. The court reasoned that the state needed the time to assure that the party had persons capable and willing to serve as electors.[216] As the dissent noted, however, since the party submitted 41,499 petition signatures to the state, it was rather unlikely that it could not muster eleven electors.[217] In an attempt to substantiate the fairness interest, one court has made the Kafkaesque suggestion that the statutes keeping a third party off the ballot actually help the third party, because they prevent unauthorized use of the party name.[218]

Ballot-access restrictions are most likely to be effective in combatting corrupt efforts to promote third-party or independent candidates as fronts for other interests—usually a major party hoping the third party will siphon votes from its principal rival. However, the disclosure provisions of FECA and similar state laws have made the state interest in combatting such practices less compelling.[219] To the extent the problem persists, restricting ballot access seems to be both an underinclusive and overinclusive solution. Ballot-access laws are underinclusive because they do not prevent outsiders from taking over a party which does qualify for the ballot—a situation exacerbated by the fact that the laws are not usually applied to the most prom-

[215] 863 F.2d 1368 (8th Cir. 1988).
[216] *Id.* at 1374–75.
[217] *Id.* at 1376 (Heaney, J., dissenting).
[218] *See* Libertarian Party v. Florida, 710 F.2d 790, 795 (11th Cir. 1983).
[219] *See supra* text accompanying notes 181–185.

ising takeover targets, the Republican and Democratic parties.[220] They are overinclusive in that they tend to penalize small parties for acts which are more likely to be perpetrated by the Republican and Democratic parties, whose members write the laws. Third parties lack the resources to worry about creating still other parties as fronts.

Thus, the role that ballot-access laws play in preventing fraud and promoting honest and fair elections is negligible. Instead, they may encourage political movements to attempt to take over parties that already have ballot status. Restrictions also encourage small parties to nominate candidates with names similar or identical to those of major party candidates, hoping to win enough votes to qualify the party automatically for the next election.[221] To the extent that ballot-access laws induce such practices, they do not prevent fraud and confusion, but may very well promote it.

In sum, ballot-access restrictions do not meaningfully advance the state interests regularly asserted to justify them. Single-member districts and informal barriers effectively limit the number of candidates that appear on the ballot in most cases. Additional ballot-access restrictions are likely to create political instability and voter confusion, and hinder state efforts to assure fair and honest elections. State interests burdening the right to vote, the first amendment right to freedom of association, and the right to equal protection of the laws must be advanced in the least restrictive manner possible. The ballot laws in effect in most states fall far short of this constitutional requirement.

While a stronger nexus between legitimate state interests and ballot restrictions might lessen the burdens these laws place on voting and associational rights, serious equal protection questions would remain. These fourteenth amendment claims alone should trigger heightened scrutiny of ballot-access laws.

C. *An Alternative Analysis: Clearing the Channels of Change*

Equal protection analysis traditionally asks if a rational basis can be found to support the legislature's distinction among groups of citizens. If a rational basis can be found, the law will

[220] *See supra* text accompanying note 207, discussing 1986 Illinois elections.
[221] *See* Jardine, *supra* note 112, at 303–04.

normally withstand judicial scrutiny. For certain "discrete and insular" racial, ethnic, and religious minorities, however, the Supreme Court has applied the more demanding strict scrutiny standard of review.[222] At first glance, third parties would seem to have little in common with such groups. Like other minorities, though, third parties have distinguishing characteristics that have excluded them from the normal give-and-take of the political process.[223] For this reason, ballot restrictions that serve to exclude third parties not only implicate rights of association, but also are subject to a pure equal protection challenge.

The Court has recognized that any effort by the legislature to exclude certain groups outright from the political process must be subject to strict scrutiny on equal protection grounds.[224] Professor Ely has termed this process "unblocking stoppages in the democratic process."[225] Laws which restrict the right of third parties and independents to appear on the ballot, or which require them to meet burdens not required of the ruling parties in order to do so, seem to be the quintessential subject of more demanding equal protection review. "Malfunction occurs when the ins are choking off the channels of political change to ensure that they will stay in and the outs will stay out, or, though no one is actually denied a voice or vote, representatives . . . are systematically disadvantaging some minority out of simple hostility."[226] It goes without saying that state legislators, who are almost all Republicans and Democrats, will always have an incentive to limit the number of challengers for their seats.[227] Third parties will typically be unable even to join coalitions to lobby for their places on the ballot, since the most they can promise a legislator in return for supporting him is that they will *not* vote for him. It is not surprising that legislators have used

---

[222] *See* United States v. Carolene Prods. Co., 304 U.S. 144, 152–53 n.4 (1938). *See also* Note, *Fairness in the Election Arena*, *supra* note 178, at 910.

[223] *See* J. ELY, DEMOCRACY AND DISTRUST 77–104 (1981); L. TRIBE, *supra* note 77, § 6-35, at 545.

[224] Kramer v. Union Free School Dist., 395 U.S. 621, 627–28 (1969).

[225] J. ELY, *supra* note 223, at 117.

[226] *Id.* at 103.

[227] *See* Note, *Access to the General Election Ballot*, *supra* note 139, at 1136 n.87 (suggesting that for this reason judicial deference may not be appropriate in ballot-access cases). Major party legislators might encourage additional challengers, however, when they believe new entrants would siphon votes from their principal rival party. Given the difficulties in making this prediction, and preventing still other rivals from entering the field once restrictions are eased, the two major parties naturally prefer a truce at the expense of all third parties.

their regulatory power over elections to attempt to stamp out political opposition from minor parties and independent candidates. Many current ballot-access laws create capricious classifications which would violate the constitutional guarantee of equal protection if strict scrutiny were applied.

Specific examples of arbitrary legislative restrictions of ballot access abound. Oklahoma's treatment of third-party candidates is particularly illustrative.[228] Oklahoma has never had more than eight candidates appear on the ballot for any statewide or United States office.[229] No more than two candidates for any single office have ever qualified as nominees of third parties; and no more than three third parties have ever appeared on the ballot in a single year, an event which has not occurred since 1936.[230] From 1944 to 1966, not a single third party qualified for the ballot.[231] Beginning in 1968, though, the American Independent Party qualified in three consecutive elections.[232] In 1974 Oklahoma raised its signature requirement for third parties to five percent of the most recent vote for President or Governor.[233] Previous law had placed third parties on the ballot with only 5000 signatures.[234] The 1974 change required approximately 52,870 signatures at the time of enactment.[235] Under the new law, the American Independent Party failed to qualify for the ballot in 1974.[236] It is hard to see this change in the Oklahoma election law as anything but an attempt to monopolize the ballot for the Republican and Democratic parties by inhibiting the American Independent Party's access to the electorate.

Oklahoma is not alone. Among the other states that have capriciously raised their signature requirements in recent years

---

[228] Oklahoma's treatment of independent candidates is considerably more reasonable. *See supra* text accompanying notes 212–214.

[229] *See* OKLAHOMA ELECTIONS, *supra* note 213; 1988 OKLAHOMA RESULTS, *supra* note 213.

[230] *See* OKLAHOMA ELECTIONS, *supra* note 213; 1988 OKLAHOMA RESULTS, *supra* note 213.

[231] *See* OKLAHOMA ELECTIONS, *supra* note 213, at D-318 to D-478.

[232] *See id.* at D-494 to D-522.

[233] 1974 Okla. Sess. Laws 153, § 1-108 (codified at OKLA. STAT. ANN. tit. 26, § 1-108 (West 1976)).

[234] OKLA. STAT. ANN. § 1-108 (West 1976).

[235] The number of Oklahomans voting for President in 1972 was 1,057,396. OKLAHOMA ELECTIONS, *supra* note 213, at D-521.

[236] *Id.* at D-536.

are Colorado,[237] Alaska,[238] Arkansas,[239] Hawaii,[240] Kansas,[241] and Maine.[242] Because no more than a handful of third parties had ever qualified for the ballot in any of these states, voter confusion does not seem to have been an imminent risk. Instead, these laws are an attempt by legislatures to limit the ballot to the Democratic and Republican parties.

Not only have legislatures been quick to erect ballot-access barriers, they have exhibited a marked tendency to ignore court holdings specifically striking down their ballot-access requirements.[243] A federal court struck down Arkansas's April deadline

---

[237] In 1989 Colorado raised its signature requirements for a third-party candidate for the legislature to appear on the ballot from 300 signatures to 1000 signatures, 1989 Colo. Sess. Laws 39, § 10 (codified at COLO. REV. STAT. § 1-4-801(b) (Supp. 1989)), even though only four third-party or independent candidates for the entire state legislature appeared on the ballot in 1988. Ballot Access News, Dec. 24, 1989, at 2, col. 1. Since the state requires just 500 signatures from an independent or third-party candidate for larger congressional districts, this law is almost certainly unconstitutional. *Id. See* Illinois Board of Elections v. Socialist Workers Party, 440 U.S. 173 (1979), discussed *supra* text accompanying notes 117–119.

[238] In 1980 Alaska raised its signature requirements for third parties seeking ballot access for offices other than President from 1000 signatures to three percent of the votes cast in the last general election (approximately 4000 signatures). 1980 Alaska Sess. Laws 100, § 138 (codified at ALASKA STAT. § 15.25.160 (1989)). In 1982 the Supreme Court of Alaska found that this requirement violated the state constitution. Vogler v. Miller, 651 P.2d 1 (1982). In 1986 the legislature lowered the requirement to one percent. 1986 Alaska Sess. Laws 85, § 26 (codified at ALASKA STAT. § 15.25.160 (1989)). Only three third parties, including one in 1978, had met the original 1000-signature requirement since statehood. *See* Letter from Richard Winger, Field Representative, Coalition for Free and Open Elections, to Brad Smith (Jan. 10, 1990) [hereinafter Winger Letter] (on file at the HARV. J. ON LEGIS.).

[239] Arkansas did not require a third party to demonstrate any level of support prior to 1971, when it began requiring signatures equal to seven percent of the last presidential or gubernatorial vote. 1971 Ark. Acts 829, § 3. This has been changed several times since and is now at three percent of the last gubernatorial vote. ARK. STAT. ANN. § 7-7-203(g) (Supp. 1989). No more than one third-party candidate has appeared on the Arkansas ballot since 1952. *See* Winger Letter, *supra* note 238.

[240] Hawaii had only two third parties qualify for the ballot during its first 10 years of statehood, even though it required no show of preliminary support. *See* Winger Letter, *supra* note 238. In 1970, however, it began requiring signatures from third parties. 1970 Haw. Sess. Laws 26, § 2 (codified at HAW. REV. STAT. § 11-62 (1988)).

[241] Kansas had only three third parties appear on the ballot between 1926 and 1966. *See* Winger Letter, *supra* note 238. Nevertheless, in 1965 it enacted a five-percent signature requirement. 1965 Kan. Sess. Laws 251, § 1 (codified at KAN. STAT. ANN. § 25-302(a) (1986)). Kansas's present requirement is two percent of the total votes in the most recent gubernatorial election. KAN. STAT. ANN. § 25-302(a) (1986). No third party has qualified under the law since 1966, although several candidates have qualified as independents and third parties have been placed on the ballot by court order. *See* Winger Letter, *supra* note 238.

[242] Maine raised its required signatures from one percent to three percent of the last gubernatorial vote in 1975 and has since raised the requirement to five percent, 1975 Me. Laws 752, § 2 (codified at ME. REV. STAT. ANN. tit. 21-A, § 303 (1989)), even though only one third party had qualified via petition since the old law was enacted in 1953. *See* Winger Letter, *supra* note 238.

[243] Note, *Fairness in the Election Arena*, *supra* note 223, at 922.

for new political parties as unconstitutional in 1977.[244] The legislature changed the deadline to May after the decision, but moved it back to January in 1987.[245] New Jersey's late-April filing deadline for independent presidential candidates was held unconstitutional in 1984,[246] yet in 1985 the legislature amended the statute to require candidates for offices other than President or Vice President to file two weeks earlier in April.[247] Although the Supreme Court specifically held as early as 1974 that a reasonable method to attain a place on the general-election ballot must exist for independent candidates as well as third-party candidates,[248] Michigan did not adopt any procedure by which an independent could qualify for the ballot until 1988.[249] In the interim years, an independent had to file a lawsuit to appear on the Michigan ballot. Ten such suits were filed from 1976 to 1984.[250] Despite the *Lubin* and *Bullock* decisions striking down mandatory filing fees,[251] four states continue to require some type of fee to appear on the ballot.[252]

---

[244] American Party v. Jernigan, 424 F. Supp. 943 (E.D. Ark. 1977).

[245] *See* 1977 Ark. Acts 888; 1987 Ark. Acts 248 (codified at ARK. STAT. ANN. § 7-7-203 (Supp. 1989)).

[246] LaRouche v. Burgio, 594 F. Supp. 614 (D.N.J. 1984).

[247] 1985 N.J. Laws 92, § 7 (codified at N.J. STAT. ANN. § 19:13-9 (1989)). New Jersey's current deadline for President and Vice President is 99 days before the general election, which would fall in late July or early August. N.J. STAT. ANN. § 19:13-9 (1989). If one interprets *Anderson v. Celebrezze* as applying to filing deadlines for third-party as well as independent candidates, 26 states have unconstitutional filing deadlines. *See* Winger Manuscript, *supra* note 46, at 2.

[248] *Storer*, 415 U.S. 724, 738 (1974).

[249] 1988 Mich. Pub. Acts 116 (codified at MICH. COMP. LAWS ANN. § 168.590 (West 1988)).

[250] Note, *Fairness in the Election Arena*, *supra* note 223, at 922.

[251] *See supra* text accompanying notes 161–169.

[252] Winger Manuscript, *supra* note 46, at 1–2. Florida charges third-party and independent candidates 10 cents to check each petition signature. Minor parties cannot waive this fee, although a random sample checking method is available if additional signatures above the minimum requirement are collected. FLA. STAT. ANN. § 99.097 (1) & (4) (West 1982). Since Florida requires a third party to submit over 180,000 signatures to appear on the 1990 state ballot, Ballot Access News, Feb. 12, 1990, at 5, this statute could result in a fee of over $18,000. Recognizing the availability of random sampling, however, the Eleventh Circuit approved this law in dicta in Libertarian Party v. Florida, 710 F.2d 790 (11th Cir. 1983), *cert. denied*, 469 U.S. 831 (1984).

North Carolina charges five cents per signature to check signatures, with no provision for waiver. N.C. GEN. STAT. § 163-96(b) (1987). This fee was over $2,000 for a candidate for state office in 1990.

Nevada has a mandatory $250 filing fee for independent candidates for President. NEV. REV. STAT. ANN. § 24-298.109 (Michie 1990). Utah also has no provision for waiving filing fees. UTAH CODE ANN. § 20-3-14 (1984). The Socialist Workers Party was denied standing to challenge this law in Hoyle v. Monson, 606 P.2d 240 (Utah 1980), on the ground that the party was not a pauper.

Despite this lengthy resume of legislative misbehavior, the Supreme Court has only twice even acknowledged that state ballot-access restrictions might relate more closely to snuffing out political competition than to meeting legitimate state objectives. On both occasions the Court has rejected the argument without reaching the merits of the claim. In *Munro v. Socialist Workers Party*, the Court specifically overturned the lower court's holding that the state's purported interest must be supported by evidence.[253] In *American Party v. White*, the Court noted plaintiff counsel's oral argument that the Texas legislature would raise the ballot-access barriers higher if any third party were to meet those at issue, but saw no reason to reach such contentions.[254]

Thus, although the Supreme Court has been hesitant to question the legislative motivations for ballot restrictions, there is ample evidence that such laws have improper purposes which violate equal protection and call for maximum judicial scrutiny.

IV. Conclusion

That ballot-access laws burden first and fourteenth amendment rights does not mean that the Supreme Court should hold all such laws unconstitutional. Though the connection between ballot-access regulations and state interests is often tenuous, the possibility exists that ballot overload could occur in rare cases. Some mechanism for regulating ballot access must remain available to the states.

The Court should not impose hard and fast rules as to exactly what that mechanism should be.[255] States ought to have the

---

[253] 479 U.S. 189, 196 (1986). *See supra* text accompanying notes 126–135.

[254] 415 U.S. 767, 772–73 n.4 (1974).

[255] Optimal mixes have been proposed. For example, a 1974 Brookings Institution report argued that "[t]he petition requirement should be no more than 1 or 2 percent of the electorate, the filing date no more than two months away from the election." D. MAZMANIAN, *supra* note 8, at 151. Since 1940, the American Civil Liberties Union has advocated a model law including petition requirements of one tenth of one percent of registered voters and automatic ballot access for parties receiving one percent of the vote in the last election. *See* AMERICAN CIVIL LIBERTIES UNION, MINORITY PARTIES ON THE BALLOT (1943).

In each of the last several Congresses, U.S. Representative John Conyers (D-Mich.) has introduced a bill closely tracking the ACLU model, but including a maximum petition requirement of 1000 signatures and a retention requirement of 20,000 votes. *See* H.R. 1582, 101st Cong., 1st Sess. (1989). The bill would only apply to candidates for federal office. The history of ballot laws indicates that even the requirements of the Conyers bill would usually be higher than necessary to address the state interests involved.

flexibility to choose among and combine the regulatory strategies available. Rigid judicial guidelines would more closely resemble a code of regulation than a constitutional principle.

Instead, the judiciary should consistently apply a true "strict scrutiny" analysis—that is, "strict scrutiny" with a "least drastic means" alternative. To be effective, this analysis requires the courts to look skeptically at asserted state interests. If there is one thing Democrats and Republicans agree on, it is a desire to keep the legislative pie divided into two large slices. Courts must carefully examine whether legislators are reacting to real threats against the public interest or only to threats against their incumbency.

To provide meaningful protection, such scrutiny cannot be limited to asking whether a ballot-access restriction makes it absolutely impossible for third parties to reach the ballot. The *Jenness* approach overlooks the burdens placed on first and fourteenth amendment rights each time a legitimate third party fails to make it onto the ballot. Furthermore, the *Jenness* approach ignores the fact that the occasional third party which does make it onto the ballot is often left without resources to continue its campaign. The *Jenness* approach of providing relief only if parties are absolutely barred from the ballot is inadequate to address third-party rights.

The Supreme Court has stated that "the right to vote freely for the candidate of one's choice is of the essence of a democratic society, and any restrictions on that right strike at the heart of representative government."[256] When state legislatures composed exclusively of Democrats and Republicans engage in tactics that place legal handicaps on all political competition, it is incumbent on the judiciary to subject the legislature's judgment to the strictest scrutiny. Only a judiciary that views ballot-access legislation with a more skeptical eye will be capable of striking a realistic balance between state interests and voter rights.

---

[256] Reynolds v. Sims, 377 U.S. 533, 555 (1964).

COMMENT

BUDGETARY TREATMENT OF FEDERAL CREDIT PROGRAMS

J. EDMUND COLLOTON, JR.*

The federal government plays a significant role in the United States credit market, particularly in the areas of housing, education, and farm financing. The basic purpose of federal credit programs is to provide credit where it would otherwise be prohibitively expensive or unavailable by subsidizing the cost of the credit. However, such subsidies are not accurately accounted for in the federal budget.

In this Comment, Mr. Colloton suggests that Congress should include the value of credit program subsidies in the budget. Mr. Colloton proposes two ways of measuring the value of these subsidies: either by the subsidy's actual cost to the government, or by the subsidy's price in the market. Mr. Colloton explores the merits of both options as well as the means of making the transition to a policy of accurate budgetary treatment of credit programs.

Given the heightened concern over federal spending in the Gramm-Rudman-Hollings[1] era, federal credit programs have received remarkably little attention. Although the size of the federal deficit and federal borrowing in the credit markets have gained much notoriety, few are aware that the federal government is the largest supplier of credit to the U.S. credit markets.[2] At the end of fiscal year 1989, federal direct and guaranteed loans outstanding totaled nearly $800 billion.[3] If loans made by government-sponsored enterprises ("GSEs") are included,[4] the federal government affected the allocation of over $1.5 trillion

---

* Associate at Davis Polk & Wardwell. A.B., Cornell University, 1983; J.D., Harvard Law School, 1990. Special thanks to Professor Howell Jackson of Harvard Law School for his helpful criticism of an earlier draft of this Comment.

[1] The Balanced Budget and Emergency Deficit Control Act of 1985, Pub. L. No. 99-177, 99 Stat. 1038 [hereinafter Gramm-Rudman-Hollings Act], set maximum deficit levels for federal spending for fiscal years 1986–1991 and provides for across-the-board spending cuts if in any year the projected deficit exceeds the maximum target level by more than a specified amount.

[2] In 1988 the federal government directly held loans with a face value 61% greater than those held by the largest U.S. commercial bank. OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT, FISCAL YEAR 1990, SPECIAL ANALYSES F-1 (1989) [hereinafter SPECIAL ANALYSIS F]. In 1988 the total amount of federal direct and guaranteed loans outstanding was more than the total amount of commercial and industrial loans outstanding. L. BRYAN, BREAKING UP THE BANK 68 (1988).

[3] Direct loans outstanding were $207 billion; loan guarantees outstanding were $588 billion. OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT, FISCAL YEAR 1991, at 229 (1990).

[4] At the end of fiscal year 1989, GSE loans outstanding totaled $763 billion. *Id.*

worth of currently outstanding credit, an amount thirty-six percent larger than the current federal budget.[5] Over the past ten years, federal credit programs have influenced the allocation of nearly one-fifth of all credit extended in the U.S. economy.[6] Despite this presence at considerable cost to the taxpayer,[7] federal credit programs have been subject to little direct scrutiny. Part of the problem lies with their current treatment in the federal budget. Cash accounting methods obscure the real costs of federal credit, leaving Congress and the public in the dark about their net value. This Comment analyzes how credit programs are included in the unified budget and explores and critiques several possibilities for reform of their budgetary treatment.

I. Background[8]

A. *The Scope of Federal Credit*

Unlike most lending in the private sector, federal credit programs operate to provide credit subsidies to borrowers.[9] For most programs, the value of what is given to a borrower is greater than the present value of what the government expects to receive in return from that borrower.[10] Federal credit is in-

---

[5] In 1989 outlays by the federal government totaled $1.14 trillion. *Id.* at 2.

[6] This figure is the average for the period 1979–1988. Special Analysis F, *supra* note 2, at F-92 (table F-22).

[7] Although the cost to the federal government of federal credit programs is approximately $20 billion per year, the economic effects of these programs may be far greater because to the extent that the supply of credit is inelastic, other borrowers are "crowded out" of the credit markets. As a result, resources may be diverted from more productive to less productive activities. *See* W. Gale, Economic Effects of Federal Credit Programs (Aug. 1989) (unpublished manuscript on file at the Harv. J. on Legis.).

[8] In the proposed budget for fiscal year 1991, the Office of Management and Budget discontinued the use of Special Analyses, including Special Analysis F, which in previous years contained detailed information on federal credit programs. As a result, information in this section is drawn from fiscal year 1990 budget documents.

[9] Even those credit programs that operate on an actuarially sound basis provide credit subsidies to borrowers. Without federal credit programs to pool risks across a large number of participants, borrowers would pay higher rates of interest.

[10] Theoretically, the loan or guarantee would not be made unless the overall benefit of providing the subsidy were greater than its overall cost. But because not all the benefits from the subsidy are fully quantifiable or flow directly to the government, the cost of the subsidy to the government is said to exceed the benefit accruing to it.

tended to compensate for credit market imperfections[11] or to achieve certain goals, such as promoting education or encouraging home ownership.[12] Credit programs make loans more affordable to borrowers and in some cases provide borrowers access to credit that they otherwise could not obtain at any interest rate.

Through credit programs, the federal government provides credit subsidies to borrowers in three forms: direct loans, guaranteed loans, and GSE loans.[13] As a direct lender, the government originates and services loans.[14] Direct loans are most justified when borrowers cannot obtain credit even with a loan guarantee. Direct loan subsidies go primarily to two groups: farmers (78%) and homeowners (18%).[15] The present value of subsidies to be provided to borrowers through direct loans made in fiscal year 1990 is estimated at $1 billion, or eight percent of total loan obligations of $12.3 billion.[16]

The bulk of government credit subsidization is in the form of guaranteed loans. A loan guarantee is a promise by the government to pay all or part of the principal and interest on a loan if the primary obligor defaults. Loan guarantees reduce the interest cost to a borrower because some or all of the risk of default is transferred from the lender to the government. Nearly all loan guarantee subsidies are provided to three groups: students (49%), homeowners (36%), and farmers (12%).[17] The present value of subsidies to be provided to borrowers through guar-

---

[11] If credit markets were perfect, borrowers of equal risk would be charged identical rates of interest. Market failure occurs when private lenders charge equally risky borrowers different rates of interest or negotiate different terms because of the borrower's size or geographic location, or as a result of taking into account factors unrelated to credit risk such as race, sex, or the type of activity of the borrower.

[12] More generally, credit programs can be considered a tool of fiscal policy.

[13] The federal government also provides credit subsidies in the form of tax-exemption of state and local bonds. Since these exemptions are "tax expenditures," this Comment does not address their budgetary treatment. Other significant federal credit subsidies include federal deposit insurance, pension insurance, crop insurance, and overseas investment insurance. But because these programs represent contingent liabilities of the federal government rather than recurring loan programs, their budgetary treatment is not addressed.

[14] B. Bosworth, A. Carron & E. Rhyne, The Economics of Federal Credit Programs 4 (1987).

[15] Office of Management and Budget, Budget of the United States Government: Fiscal Year 1990, at 6-23 (1989).

[16] Special Analysis F, *supra* note 2, at F-13 (table F-6), F-42 (table F-12). This calculation of the subsidy does not include write-offs for defaults, which in 1988 for direct loan programs were $2.5 billion. *Id.* at F-44 (table F-14).

[17] Office of Management and Budget, Budget of the United States Government: Fiscal Year 1990, at 6-23 (1989).

antees made in fiscal year 1990 is estimated at $8.3 billion, or seven percent of total loan guarantees of $111.7 billion.[18]

The federal government also participates significantly in the credit markets through GSEs. GSEs are government-chartered entities that enjoy special benefits from their affiliation with the federal government. GSEs are off-budget agencies[19] because there is no legal obligation to support them in the event of insolvency. For the most part, GSEs are supported only by a "moral" obligation on the part of the government to provide assistance in the event of financial difficulty. Though implicit, this obligation carries significant weight in the credit markets, and, as a result, most GSEs can borrow at interest rates only slightly above the treasury borrowing rate. In some cases, GSE obligations are supported by statute.[20]

Since 1975 loans by GSEs have grown at an annual rate of nearly nineteen percent.[21] Generally, GSEs either act as credit intermediaries, helping create secondary markets,[22] or lend to special groups at a discount from the market rate of interest.[23] GSE programs primarily support agriculture, education, and housing.[24] The government estimates that GSEs will obligate loans of $422.4 billion in 1990.[25] Since most GSEs operate on an actuarially sound basis, arguably the direct cost to the government of these programs is zero.[26]

---

[18] SPECIAL ANALYSIS F, *supra* note 2, at F-18 (table F-7), F-43 (table F-13). This calculation of the subsidy does not include terminations for defaults, which in 1988 for guaranteed loan programs were $11.2 billion. *Id.* at F-44 (table F-14).

[19] The Farm Credit System Financial Assistance Corporation ("FAC") is an exception to this rule. FAC was created in 1988 to help bail out the ailing Farm Credit System, which lends to farms and farmer-owned cooperatives. 12 U.S.C. §§ 2278b, 2278b-1 (1988).

[20] For example, the Federal National Mortgage Association, a GSE, is authorized to borrow $2.25 billion from the Treasury Department in the event of financial problems. D. IPPOLITO, HIDDEN SPENDING 38 (1984). Also, the government is obligated to "back up" interest payments on obligations issued by the Resolution Funding Corporation, a GSE established to assist in the thrift bailout crisis. Financial Institutions Reform, Recovery, and Enforcement Act of 1989, Pub. L. No. 101-73, § 511(f)(2)(E)(i), 103 Stat. 183.

[21] SPECIAL ANALYSIS F, *supra* note 2, at F-32.

[22] The Federal National Mortgage Association, for example, purchases home mortgages from lenders and issues mortgage-backed securities that guarantee payment of principal and interest. *See* 12 U.S.C. § 1717 (1988). This has helped create a large secondary market for home mortgages and has lowered interest costs to borrowers.

[23] The Farm Credit System, for example, lends to farmers and to farmer-owned cooperatives.

[24] SPECIAL ANALYSIS F, *supra* note 2, at F-23.

[25] *Id.* (table F-9).

[26] Most GSEs, however, do enjoy at least indirect federal subsidization, including exemption of corporate earnings from federal income tax, exemption of interest income of investors from state and local tax, and exemption from SEC registration and various state banking laws. SPECIAL ANALYSIS F, *supra* note 2, at F-22 (table F-8). In addition,

B. *Current Treatment of Credit Programs in the Budget: Something for Nothing*

The unified budget is an annual cash flow document that is not well suited to measuring the costs of federal credit programs. When a loan of $1 is made, it is entered in the budget as an outlay of $1, equivalent to an expenditure of $1, even though repayment of the loan is anticipated.[27] Initially, then, the cost of extending credit is overstated in the budget. This accounting approach favors existing credit programs because, once initial outlays are made, annual outlays are required only to the extent of *net* cash requirements. Repayment of loans (both principal and interest) and the sale of loan (and other agency) assets to the public are credited as negative outlays.[28] As a result, old loan programs can be continued at minimal perceived cost.

The interest rate subsidy, which is the lower interest rate paid by the borrower as a result of the federal loan or guarantee, accrues over the life of the loan or guarantee and varies considerably depending on the credit program.[29] This subsidy is not included in the budget when the loan or guarantee is made. Interest costs affect the program's ability to make new loans or guarantees, but are not explicitly budgeted when credit is extended.

At origination guaranteed loans require no budget authority at all.[30] Loan guarantees are considered only contingent liabilities and require budget authority and outlays only upon the

---

although there may be no cost to the government, the economic costs of GSE programs may be significiant because other borrowers may be crowded out of the credit markets. *See supra* note 7.

[27] For the well-established loan programs, the Congressional Budget Office ("CBO") has estimated that nearly 95% of the outstanding loans will eventually be repaid. D. IPPOLITO, HIDDEN SPENDING 5–7 (1984).

[28] Not all sales of loan assets affect the deficit as calculated for Gramm-Rudman-Hollings Act purposes. Asset sales not "routine and ongoing" in fiscal year 1986 or approved in legislation enacted before September 18, 1987, do not reduce the Gramm-Rudman-Hollings Act deficit calculation. SPECIAL ANALYSIS F, *supra* note 2, at F-51.

[29] For direct loans, the total value of the subsidy (excluding defaults) ranges from 0.1% to 71.2% of the total value of the loan. *Id.* at F-42 (table F-12). For guarantees, the total value of the subsidy (excluding defaults) ranges from 0.7% to 32.3% of the amount guaranteed. *Id.* at F-43 (table F-13). Much of the value of subsidies comes in the form of reduced interest rates, but subsidies also include benefits arising from longer maturities, deferred interest, extended grace periods, the waiver or reduction of loan fees, and loans larger than those available in the private sector. *Id.* at F-39.

[30] Loan guarantees are excluded from the definition of budget authority under §§ 3(a)(2) and 401(c)(2) of the Budget Act of 1974. Pub. L. No. 93-344, 88 Stat. 297. Early versions of the Budget Act would have included administrative costs and anticipated default losses as present outlays, but these provisions were not included in the final legislation. CONGRESSIONAL BUDGET OFFICE, LOAN GUARANTEES: CURRENT CONCERNS AND ALTERNATIVES FOR CONTROL 8 (1978).

default of primary obligors. Moreover, guarantee fees received from borrowers are recorded immediately. Thus, guarantee programs often initially run net surpluses. This budgetary practice favors guarantee programs over direct lending programs, even though their subsidy costs may be identical.[31] For loans and guarantees made in 1990 it is estimated that subsidies provided under loan guarantee programs will be eight times greater than those furnished under direct loan programs.[32] Between 1987 and 1990, direct loans outstanding are projected to decrease at four percent annually, while guaranteed loans outstanding are projected to increase five percent annually.[33]

In both direct and guaranteed loan programs, default costs are not recognized until well after the loans or guarantees have been made. Although often expected, default costs are not recorded as outlays until actual defaults occur. Even then, often loans are refinanced to avoid currently incurring the cost of the default. (Typically, the loan's maturity is extended or interest payments are reduced.) The refinanced loan is then substituted for the defaulted loan, deferring the recognition of loss. Some defaults are carried forward as currently outstanding loans, even when it is clear that they will not be repaid. For example, as late as 1972, World War I loans to U.S. Allies remained on the government's books despite being long in default and despite their non-recognition by the borrowers.[34] This practice is attractive because, to the extent that old loans are written off, the ability to make new loans is diminished. In addition, when guaranteed loans go into default, they are often purchased from private lenders to avoid recording the cost of default currently.[35]

Ideally, a budget facilitates the allocation of scarce resources among competing programs and alternative uses. When budgetary practices reflect the cost of old rather than new loans and guarantees, the allocation process becomes distorted.

---

[31] Only since the Federal Financing Bank ("FFB") was put "on-budget" in 1987 have loan guarantee programs been the more preferred means of providing credit subsidies. Prior to 1987, direct loans could be sold to the FFB, then an off-budget agency, freeing up funds for new loans. After the FFB was put on-budget in 1987, *see* Gramm-Rudman-Hollings Act, *supra* note 1, § 214, loans sold to it did not create a negative outlay. B. Bosworth, A. Carron & E. Rhyne, *supra* note 14, at 155.

[32] Special Analysis F, *supra* note 2, F-42 (table F-12), F-43 (table F-13).

[33] *Id.* at F-13 (table F-6), F-19, F-92 (table F-22).

[34] D. Larkins, 300 Billion in Loans 44 (1972). As late as 1987, loans made by the Export-Import Bank to Cuba in the 1950's were still held on the agency's books at face value. W. Gale, The Hybrid Plan: A Proposal for Federal Credit Reform 7 (Jan. 1990) (unpublished manuscript on file at the Harv. J. on Legis.).

[35] While this practice does result in an outlay, the agency's balance sheet is distorted to the extent that it is expected that these loans will not be repaid.

C. *Congressional Controls*

In 1980 the President for the first time included an explicit budget for credit programs in his annual budget submission to Congress.[36] The President requested that Congress set annual volume limits (credit authority) on new direct loans and guarantee commitments, and Congress has generally adopted this approach.[37]

Since budget authority and outlays are required only if a loan program runs a net deficit during the year, Congress uses credit authority to control the volume of new loans and guarantees on a gross basis. Currently, the annual budget resolution includes a credit budget, and appropriation bills do contain some limitations on credit authority. In 1988 thirty-nine percent of direct loans and fifty-seven percent of guaranteed loan commitments were subject to credit authority limitations.[38] But these limitations are ineffective. Besides not applying to all credit programs, the limitations are often set too high to have any impact on program activity.[39] Additionally, by controlling programs with credit authority alone, Congress fails to distinguish among programs based on differences in interest rate subsidies, default rates, or payment terms. While credit authority limitations do control the amount of new loans that are made, they do not control the interest rate at which loans are made, mandate minimum credit standards, or specify the term over which loans must be made.

The Gramm-Rudman-Hollings Act, which in 1985 amended the Budget Act of 1974, moved the Federal Financing Bank ("FFB") on-budget, an important ramification for credit program budgeting.[40] The FFB was created in 1973 as an agency of the Treasury Department to coordinate agency borrowing. Prior to the Gramm-Rudman-Hollings Act, agencies could sell loan as-

---

[36] CONGRESSIONAL BUDGET OFFICE, EXPLICIT BUDGET FOR FEDERAL CREDIT PROGRAMS: FEDERAL CREDIT ACTIVITY: AN ANALYSIS OF THE PRESIDENT'S CREDIT BUDGET FOR 1981, at 8 (Feb. 1980).

[37] In December 1985 The Congressional Budget Act was amended to require that the credit budget be included in the annual congressional budget resolution. Gramm-Rudman-Hollings Act, *supra* note 1, § 212. The Gramm-Rudman-Hollings Act also requires functional allocations for direct loan obligations and loan guarantee commitments. SPECIAL ANALYSIS F, *supra* note 2, at F-8, F-9.

[38] *Id.* at F-8 (table F-3).

[39] *Id.* at F-10.

[40] *See supra* note 31. The FFB is authorized to (1) purchase loan assets from agencies, (2) disburse loans to borrowers (when loans are 100% guaranteed by the agency), and (3) buy debt from agencies that are authorized to borrow from the public. *See* A. WILDAVSKY, THE NEW POLITICS OF THE BUDGETARY PROCESS 125–33 (1988).

sets to the FFB and the sale was recorded as a negative outlay. Agencies could then make new loans or guarantees without additional budget authority or outlays. Under the Gramm-Rudman-Hollings Act, the sale of loan assets to the FFB does not create the ability to make new loans or guarantees. If the FFB is kept on-budget, this will constitute an important reform.

Aside from this one key reform, federal credit programs have not been subject to the discipline of the Gramm-Rudman-Hollings Act. Credit programs are inadequately included in the Gramm-Rudman-Hollings Act sequestration process.[41] If sequestration occurs, credit authority is reduced by the same percentage that budget authority for spending on non-defense programs is reduced.[42] But since credit authority levels are set too high or are nonexistent, federal credit programs escape much of the heat of the Gramm-Rudman-Hollings Act.

Volume limits are inadequate to control credit programs. Credit authority does not apply to the majority of programs, affects only net outlays, and does not differentiate among programs based on differences in interest costs, default rates, or terms of payment. Reform is needed to identify, record, and control the cost of federal credit when the decision to extend credit is made.[43] For budgetary purposes the cost—not the amount—of federal credit needs to be highlighted.

II. Proposals for Reform

The Office of Management and Budget ("OMB"),[44] the Congressional Budget Office ("CBO"),[45] the Senate Budget

---

[41] Sequestration describes the procedure used to reduce spending across the board "automatically" to meet the deficit targets required under the Gramm-Rudman-Hollings Act if in any year the projected deficit exceeds the deficit target by more than a specified amount. *See id.* at 241–49.

[42] Gramm-Rudman-Hollings Act, *supra* note 1, § 251.

[43] *See* Office of Management and Budget, Budget of the United States Government: Fiscal Year 1990, at 6-27 (1989).

[44] Office of Management and Budget, Budget of the United States Government: Fiscal Year 1990, pt. 6 (1989); Office of Management and Budget, Budget of the United States Government: Fiscal Year 1991, at 243–46 (1990).

[45] Congressional Budget Office, New Approaches to the Budgetary Treatment of Federal Credit Assistance (Mar. 1984); Congressional Budget Office, Reforming the Budgetary Treatment of Federal Credit Programs (draft report July 1988).

Committee,[46] and the General Accounting Office ("GAO")[47] each have proposed reform of the budgetary treatment of federal credit programs. All have recommended that the budget include the value of credit subsidies[48] provided to borrowers as an outlay in the year in which the loan or guarantee is made. Policymakers disagree over the best approach to measure credit subsidies. There are two methods of measurement: (1) their "cost" to the government, or (2) the difference in "price" between the government loan or guarantee and a comparable loan or guarantee in the market.

A. *Cost to the Government*

With the cost approach, the amount of the subsidy is the difference between the amount loaned out and the present value of the stream of anticipated payments from the borrower (less default costs and administrative costs). For guarantees, the amount of the subsidy is the difference between the fee paid by the borrower for the guarantee and the sum of default costs and administrative costs, discounted to present value.

In calculating present value, the GAO Study suggests using a discount rate equal to the coupon rate paid on treasury bonds of a maturity similar to that of the loan.[49] This approach should be modified in three respects. First, the discount rate should reflect the borrowing cost incurred by the agency, not the Treasury. The discount rate should include the one-eighth of a percentage point that agencies are charged by the FFB to defray FFB administrative costs. Second, the discount rate should reflect the *yield* on treasury bonds, not their coupon rate. Treasury securities are often sold at a discount or at a premium, rather than at par. A discount rate using yield-to-maturity more accurately reflects the government's actual cost of borrowing. Third, the discount rate should be computed net of federal income tax.[50] Interest payments made on treasury securities are taxable

---

[46] *See*, S. REP. No. 40, 100th Cong., 1st Sess., at 119 (1987).

[47] GENERAL ACCOUNTING OFFICE, BUDGETARY TREATMENT OF FEDERAL CREDIT PROGRAMS, AFMD-89-42 (April 1989) [hereinafter GAO STUDY].

[48] A credit subsidy is the benefit that accrues to the borrower as a result of receiving federal, rather than private-sector, credit. *See supra* Part I(A).

[49] GAO STUDY, *supra* note 47, at 14.

[50] D. LARKINS, *supra* note 34, at 35.

as ordinary income. Thus the cost of borrowing is the treasury yield minus what is recouped through income taxes.

The GAO Study would exclude agency administrative expenses from the estimate of the subsidy. They are excluded "because of the complexities of measuring and allocating prospective administrative costs."[51] As a result, the budget would understate the credit subsidy. If prospective administrative costs cannot be determined, an historical average, adjusted for inflation, could be used. In the event that administrative costs cannot be isolated for any particular credit program, the cost per loan or guarantee dollar disbursed in a comparable program could be substituted in its place. It is a mistake to ignore these costs completely.

As the GAO Study suggests,[52] the stream of anticipated payments from the borrower should be reduced by the amount expected to be lost as a result of defaults. With established credit programs, agencies can predict default rates with a high degree of accuracy.[53] In the case of new loan or guarantee programs, default costs may be more difficult to calculate. An independent body could be established to provide credit programs with expertise in estimating prospective default losses. Because an agency may be tempted to understate default rates to reduce a credit program's estimated cost in order to increase support for the program, an independent body would help avoid agency manipulation of prospective default rates.

B. *Price in the Market*

Alternatively, the government could use a price approach to calculate the value of the subsidy provided to the borrower. The price approach measures the difference between what a private lender would charge the borrower for the loan or guarantee and what the government in fact charges the borrower. Using this approach, the subsidy can be calculated in one of two ways. The first method is the "market" method. As soon as the loan is disbursed or the guarantee is made, the agency would sell the loan on the open market or reinsure the loan guarantee in the

[51] GAO STUDY, *supra* note 47, at 5.
[52] *Id.*
[53] This assumes that economic conditions can be predicted and credit criteria remain constant.

market. For a direct loan, the subsidy is the difference between the amount received by the government in the loan sale and the amount loaned to the borrower. For a guaranteed loan, the subsidy is the difference between the cost of reinsurance and the guarantee fee paid by the borrower.[54] Under this method, the government would sell loans without recourse in order to determine the value of the subsidy accurately. By contrast, the government could hold the loan or guarantee and estimate the subsidy based on the difference between the cost of the government loan or guarantee and that of a "comparable" loan or guarantee in the private sector.

There are several problems with each of the price approaches. Under the comparable loan approach, estimating the interest rate that the borrower would have been charged in the private sector for a comparable loan or guarantee could prove to be difficult. With hundreds of loan programs in existence[55] and considerable differences in risk among borrowers, the difficulties of collecting accurate data might be insurmountable. In addition, since some federal credit programs provide loans to borrowers who could not otherwise borrow at any interest rate, some loans could not be priced at all.

The market method suffers from serious shortcomings as well. Congress created federal credit programs because the private sector had failed to price some types of credit accurately, had not treated equally situated borrowers alike, and at times had refused to lend funds at all. Consequently, at least with respect to federal credit beneficiaries, the market is an imperfect mechanism to rely on when pricing loans or guarantees. For some loans, a secondary market (or the availability of reinsurance) is virtually non-existent. Although the government has successfully created a viable secondary residential mortgage market through GSEs, presently this is the exception rather than the rule. Many loans would have to be sold at a deep discount from what it would cost the government to service them. For many guarantee programs, the cost of reinsurance would be prohibitive. The problem stems from a lack of information (loan documentation by federal credit programs is notoriously poor) and lack of standardization. The OMB suggests that market sales

---

[54] For both direct and guaranteed loans, administrative costs would also have to be deducted to determine the subsidy accurately.

[55] D. LARKINS, *supra* note 34, at 36.

would encourage credit programs to improve "loan origination and documentation."[56] In the meantime, the market method would seriously overestimate the actual subsidy because many loans would have to be sold at a significant discount.

C. *Price Versus Cost*

The price approach would overstate the credit subsidy provided to borrowers. The subsidy would be overstated because interest rates charged by private lenders include a margin for profit as well as costs incurred because of federal and state regulation. The subsidy would also be overstated if loan assets are sold at a discount because of poor loan documentation and lack of standardization.

Aside from these difficulties, those advocating the price approach (the CBO and OMB) have missed a key point. The purpose of estimating credit subsidies is neither to measure how much borrowers gain under credit programs, nor to measure the effect of federal credit on the credit markets or on the economy. Rather, the purpose of calculating subsidies is to estimate the cost of federal credit so that it can be compared with spending on other federal programs. The federal budget is not an attempt to quantify the benefits arising as a result of federal spending. The price approach mistakenly measures the benefits—not the costs—as a result of federal credit programs. Although superior to current budgetary practices, use of the price approach is misplaced in the budgetary context.

III. Moving to Subsidy Accounting

In order to adopt a subsidy approach, there must be a means of smoothly making the transition from the current budgetary treatment of credit programs to the present value subsidy approach. All of the government proposals for reform suggest creating agency credit "subsidy accounts," to which Congress would annually appropriate the estimated present value of credit subsidies resulting from loans and guarantees made that year. Subsidy accounts, of course, would be placed on-budget to reflect the true cost of federal credit in the budget.

---

[56] Special Analysis F, *supra* note 2, at F-48.

For direct loans, parallel credit "financing accounts" would be established to finance that portion of the direct loan that will eventually be repaid. For direct loans and guarantees, parallel credit financing accounts would be used to pay for defaults. The estimated costs of defaults would be included in the credit subsidy account initially, but would be transferred to the credit financing account to pay for defaults as they occur.[57] When an agency made a loan, it would draw funds from both the subsidy and finance accounts. Repayments, as they occur, would restore the finance accounts to a net zero position.

This accounting approach highlights the actual cost of federal credit programs, provided that default costs are accurately assessed and that payments made because of defaults are carefully scrutinized. Agencies must not be allowed to finance defaults by drawing from financing accounts, without having first transferred funds from subsidy accounts. While close audits of the accounts will give agencies greater incentive to accurately assess the probability of defaults when the agency first extends credit, an independent appraisal of projected defaults will enhance credit budget accuracy. By isolating default costs at the outset, this approach would enhance the incentive to improve collection efforts, as default costs would have to be drawn from subsidy appropriations alone.

The financing accounts should be placed off-budget. Placing the financing accounts on-budget would continue to overstate both outlays and the federal deficit (because, by definition, financing accounts include only amounts that will be repaid). While funds received from the repayment of old loans would to some degree offset this overstatement of outlays, keeping repayments and the financing accounts on-budget will only perpetuate the bias in favor of existing loan programs.[58]

There is one danger in placing the financing accounts off-budget. If this were done, there would be tremendous incentive to underestimate credit subsidies and fund some credit subsidies (notably defaults) out of the off-budget financing accounts. Cre-

---

[57] GAO STUDY, *supra* note 47, at 10. Interest rate subsidies would be handled in the same manner as default costs. As interest accrues over the life of the loan, that portion of the interest cost that is federally subsidized would be transferred from the subsidy account to the financing account.

[58] In addition, even in the case of existing loan programs, if the volume of loans or guarantees is altered, this approach is no longer deficit-neutral because repayments will no longer offset outlays.

ation of an independent body to audit closely credit program accounts and estimate rates of default avoids this problem.

IV. Conclusion

Sun Tzu once said, "To subdue the enemy without fighting is the acme of skill."[59] Advocates of federal credit programs have demonstrated considerable skill in securing substantial federal expenditures while avoiding the recent trench warfare of the budgetary process. However, federal credit programs affect the allocation of available credit at considerable cost to the government and to the economy. Their true cost must be accurately reflected in the federal budget so that credit program expenditures can be compared on a dollar-for-dollar basis with other federal spending. Only then can budget choices be made intelligently. Credit budget reform will also shed light on the policy choices being made in the credit budget process. And sunlight, as Louis Brandeis said, is the best of disinfectants.[60]

Budgeting the present value of subsidies given to borrowers accurately depicts the true cost of extending federal credit. Agency budgets should include subsidy costs in the year in which the loan or guarantee is made. While the cost approach is the preferable means of estimating credit subsidies and budgeting the cost of credit programs, either the cost to government or the price in the market approach would be a vast improvement over current budgetary practices.

---

[59] Sun Tzu, The Art of War 77 (S. Griffith trans. 1963).
[60] L. Brandeis, Other People's Money 62 (1933).

APPENDIX

Federal Participation in Domestic Credit Markets
Net Loans
(billions of dollars)

	1986	*1987*	*1988*	*1989*	*1990*
Net Direct Loans[61]	11.2	−19.0	−13.4	−14.6	−2.3
Net Loan Guarantees[62]	34.6	60.4	40.3	41.7	43.1
Net GSE Loans[63]	83.3	107.8	82.5	101.5	116.3
Total Net Federal and Federally Assisted Lending	129.1	149.2	109.4	128.6	157.1
Total Funds Loaned in Domestic Credit Markets[64]	904.3	727.9	745.5	N.A.	N.A.
Federal and Federally Assisted Lending as a Percentage of Total Credit	14.3%	20.5%	14.7%	N.A.	N.A.

N.A. = not available

[61] SPECIAL ANALYSIS F, *supra* note 2, at F-92 (table F-22); OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT: FISCAL YEAR 1991, at 232 (1990) (1990 figures are government estimates).

[62] SPECIAL ANALYSIS F, *supra* note 2, at F-92 (table F-22); OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT: FISCAL YEAR 1991, at 232 (1990) (1990 figures are government estimates).

[63] SPECIAL ANALYSIS F, *supra* note 2, at F-92 (table F-22); OFFICE OF MANAGEMENT AND BUDGET, BUDGET OF THE UNITED STATES GOVERNMENT: FISCAL YEAR 1991, at A-141 (1990) (1989 and 1990 figures are government estimates).

[64] *Id.*

RECENT DEVELOPMENTS

THE CLEAN AIR ACT AMENDMENTS OF 1990 AND THE USE OF MARKET FORCES TO CONTROL SULFUR DIOXIDE EMISSIONS

The Clean Air Act[1] needed rejuvenation. After two decades of attempts to improve air quality under the Act, at least one study indicated that illnesses and premature deaths resulting from breathing polluted air cost $40–50 billion annually.[2] Also, the Environmental Protection Agency ("EPA") reported that in 1989, 96 areas exceeded its ozone air quality standard and 41 areas violated its carbon monoxide standard.[3] Although those numbers were down from 101 and 44 respectively in 1988, they still indicated the widespread failure of the pre-existing law.[4]

Whether George Bush responded to these and similar studies out of environmental concern or political expediency, his promise to promote environmental legislation in Congress played a significant role in enabling him to win the votes of conservative and moderate Democrats in 1988.[5] Although his measures were not universally praised by environmentalists or congressional Democrats, his submission of proposed Clean Air Act Amendments to Congress in July 1989[6] was a refreshing change from the Reagan doldrums. Doubters in Congress and in environmental groups were pleased that a President had, for the first time in eight years, initiated the political momentum behind environmental legislation. Congressional hope for clean air legislation also was strong, as evidenced by the support of members of Congress in key positions, including Senate Majority Leader George Mitchell (D-Me.), Chairman of the Senate Environmental and Public Works Committee, Quentin N. Burdick (D-N.D.), and Chairman of the House Energy and Commerce Committee,

---

1 42 U.S.C. §§ 7401–7642 (1988).

2 *Annual Health Costs of Air Pollution Reach $50 Billion, Lung Association Says*, Current Dev., Env't Rep. (BNA) 1648 (Jan. 26, 1990).

3 *Ozone Standard Violated in 96 Areas; 41 Fail on Carbon Monoxide, EPA Says*, Current Dev., Env't Rep. (BNA) 815 (Aug. 24, 1990).

4 *See generally* H.R. REP. No. 490, 101st Cong., 2d Sess., pt. 1 (1990); S. REP. No. 228, 101st Cong., 1st Sess. (1989).

5 Throughout the campaign, Bush portrayed himself as an ardent environmentalist and attacked Massachusetts Governor and Democratic presidential candidate Michael Dukakis on the condition of Boston Harbor. *See, e.g.*, N.Y. Times, Apr. 17, 1990, at B10, col. 1.

6 N.Y. Times, July 22, 1989, at 1, col. 4.

John Dingell (D-Mich.).[7] This shared executive and legislative interest resulted in the Clean Air Act Amendments of 1990, signed into law on November 15, 1990.[8]

One of the most controversial titles in the Amendments—where the commonality of interests among members of Congress and between Congress and the President was most strained—regulates acid rain. Acid rain (technically acid deposition) is the formation of sulfuric and nitric acid in the atmosphere from the emissions of sulfur dioxide and nitrogen oxides. But scientists disagree about the damage caused by sulfur dioxide emissions.[9] These differences provide fodder for opponents of regulation. Further, more so than on any other environmental issue, the lines separating proponents and opponents of acid rain regulation are geographical.[10] The vehemence with which regional spokespersons have argued their positions, and the legislative decision to resolve those arguments by use of economic incentives, make the sulfur dioxide regulations interesting and worthy of further inquiry.

Accordingly, the analysis in this Essay concentrates on the acid rain title. First, however, Part I highlights the other major provisions of the Amendments as passed by the Senate on April 3, 1990,[11] emphasizing the legislation's broad strategies to achieve emissions reductions. Part II describes the acid rain provisions. Part III critically examines some of these provisions, concentrating on the hotly debated program of using market forces to control sulfur dioxide (and thus, acid rain) efficiently.

I. Provisions of the Amendments

The President brought companion bills to amend the Clean Air Act to the Senate and House on July 21, 1989.[12] The Senate

---

[7] Senator Burdick, for example, stated that clean air legislation was "the number one priority of the committee during the 101st Congress." *Senate Launches Clean Air Legislation*, PUB. UTIL. FORT., Apr. 13, 1989, at 38.

[8] N.Y. Times, Nov. 16, 1990, at A28, col. 1.

[9] *See infra* notes 80–87 and accompanying text.

[10] *See infra* notes 88–98 and accompanying text.

[11] References to the Amendments are to the April 3, 1990, Senate version. For practical reasons this Essay had to rely on an earlier version than that finally sent to the President. The final version is substantially similar to the Senate version; indeed, the conference committee adopted the Senate acid rain title, with only minor amendments.

[12] *Acid Rain and the Bush Clean Air Bill*, PUB. UTIL. FORT., Aug. 31, 1989, at 30.

bill (S. 1490)[13] was introduced on August 3, 1989, by Senator John H. Chaffee (R-R.I.) and twenty-four others, and hearings began in the Environmental Protection Subcommittee of the Committee on the Environment and Public Works on September 28, 1989.[14] However, by the time it emerged from the Committee, it had been partially replaced by other bills; its remaining portions were subsumed in a bill (S. 1630)[15] introduced by Senator Max Baucus (D-Mont.). S. 1630 passed the Senate on April 3, 1990.[16]

The House bill (H.R. 3030)[17] was introduced by Representative Dingell on July 27, 1989, and ultimately was incorporated into S. 1630 on May 23, 1990.[18] A House-Senate conference convened on July 13, 1989, to reconcile the two versions of S. 1630.[19] The conference report was accepted by the House (401-25) on October 26, 1990, and by the Senate (89-10) on October 27, 1990.[20] President Bush signed S. 1630 into law on November 15, 1990.[21]

The heart of the Senate bill is found in the first four titles, which address air quality standards, motor vehicle emissions, toxic air pollutants, and acid rain.[22] The other seven titles deal primarily with enforcement, housekeeping, or special interests.[23]

A. *Title I: Ambient Air Quality Standards*

Under pre-existing law, the Administrator of the EPA (the "Administrator") must promulgate national ambient air quality standards ("NAAQS") every five years. Each state must attain those NAAQS through its own State Implementation Plan ("SIP").[24] NAAQS are designed primarily to combat smog. The major pollutants regulated by the bill are sulfur dioxide, parti-

---

[13] S. 1490, 101st Cong., 1st Sess. (1989).
[14] Cong. Index (CCH) 21,025 (1989).
[15] S. 1630, 101st Cong., 1st Sess. (1989).
[16] *Id.* at 20,510, 21,026.
[17] H.R. 3030, 101st Cong., 1st Sess. (1989).
[18] Cong. Index (CCH) 34,513, 35,042 (1990).
[19] *Id.* at 20,510.
[20] 136 CONG. REC. H12,943–44 (daily ed. Oct. 26, 1990); 136 CONG. REC. S17,434 (daily ed. Oct. 27, 1990).
[21] N.Y. Times, Nov. 16, 1990, at A28, col. 1.
[22] S. 1630, 101st Cong., 2d Sess. tit. I–IV, 136 CONG. REC. S4364–432 (daily ed. Apr. 18, 1990).
[23] *Id.* §§ 501–1103, 136 CONG. REC. at S4432–448.
[24] 42 U.S.C. §§ 7409–7410 (1988).

culate matter, carbon monoxide, ozone, nitrogen dioxide, and lead.[25]

Title I of the Amendments sets up an improved framework for determining and enforcing NAAQS. It requires the governor of each state to designate all areas in the state as (1) non-attainment areas (which have failed to meet the NAAQS), (2) attainment areas, or (3) unclassifiable areas, for each pollutant.[26] Non-attainment areas for primary NAAQS must reach attainment within five years from the date of designation.[27]

Twenty-four months after the promulgation of a new set of NAAQS, states must submit SIPs that articulate enforceable limitations on emissions, schedules for compliance, monitoring provisions, and a fee program to pay for implementation.[28] If a state submits an unsatisfactory SIP, or fails to submit one, the Administrator must promulgate a Federal Implementation Plan ("FIP") for the area.[29] In addition, if a state fails to submit an adequate SIP or fails to enforce a SIP requirement (and is not making reasonable efforts to cure the failure), the Administrator must take punitive action against the state.[30]

Title I emphasizes the reduction of ozone pollution. Ozone non-attainment areas are divided into four categories—moderate, serious, severe, and extreme—based on the percentage by which they exceed the NAAQS.[31] Each category of ozone non-attainment is assigned specific required implementation programs that build upon the lower categories' requirements. For example, severe areas must revise their vehicle inspection programs, require vapor recovery systems at large gas stations, revise SIPs to attain percentage reductions of either of the two main ozone producing pollutants (volatile organic compounds and nitrogen oxides), and require employers of 100 persons or

---

[25] *See* 40 C.F.R. §§ 50.4–50.12 (1989).

[26] S. 1630, 101st Cong., 2d Sess. § 101(a), 136 CONG. REC. S4364 (daily ed. Apr. 18, 1989).

[27] *Id.* § 106(d)(1)(A), 136 CONG. REC. at S4370. The five-year limit is extendable to 10 years as long as the Administrator determines that the clean-up is progressing "as expeditiously as practicable." *Id.*

[28] *Id.* § 104, 136 CONG. REC. at S4366.

[29] *Id.* § 104(b), 136 CONG. REC. at S4367.

[30] *Id.* § 106(g)(3), 136 CONG. REC. at S4371. The Administrator's options are to prohibit construction or modification of any source of the relevant pollutant which may emit 100 tons annually in non-attainment areas; to prohibit grants of federal highway funds unless such funds will improve air quality; to withhold all or part of the states' air pollution grants awardable under Section 105 of the Act; or to prohibit the provision of new drinking water service in the non-attainment area.

[31] *Id.* § 107, 136 CONG. REC. at S4373.

more either to increase average occupancy per vehicle in commuting trips, or to spend at least as much on the attempt as on providing employee parking spaces.[32]

B. *Title II: Mobile Sources*

Title II regulates vehicle emissions of many of the same pollutants as Title I. However, where Title I limits the ambient air concentration of a pollutant, Title II limits actual emissions. The boldest provisions of Title II tighten tailpipe emissions standards and encourage or require the use of alternative fuels.

The first-round tailpipe emissions standards are to be phased in on all automobiles by the 1995 model year.[33] (The bill also sets forth specific tailpipe standards for trucks and buses.)[34] A second round of tighter standards could be triggered for model year 2004 or 2007 depending on the severity of the Title I classification and the ambient ozone levels at designated years.[35] The Administrator must grant transferable credits to an automaker whose cars have a lower average emission than required.[36] These credits may be sold by one automaker to another for cash. This permits some specialization so that automakers who can achieve the reductions most cheaply have an incentive to compensate for others.

The second major portion of Title II requires heightened use of "clean alternative fuels," which include methanol and methanol blends, ethanol, reformulated gasoline, and natural gas.[37] Beginning in 1992, in ozone non-attainment areas, all gasoline-operated vehicles must use reformulated gasoline that meets the Administrator's specifications.[38] Furthermore, fleet vehicles

---

[32] *Id.*, 136 CONG. REC. at S4373–75. Section 107 also establishes a Northeast ozone transport region, comprising the coastal states from Virginia to Maine. These states would have to apply some ozone reduction measures, regardless of their attainment status. *Id.*, 136 CONG. REC. at S4377. Section 107 also requires specific action in carbon monoxide and particulate matter non-attainment areas. For example, "serious" carbon monoxide non-attainment areas must institute carpooling incentives and other transportation control measures. *Id.* § 108, 136 CONG. REC. at S4379. The thrust of the particulate matter section is to control agricultural and forestry burning and to encourage the use of "best available control technology," such as EPA-approved wood-burning stoves. *Id.* § 109(a), 136 CONG. REC. at S4380.

[33] *Id.* § 201(a), 136 CONG. REC. at S4385.

[34] *Id.*, 136 CONG. REC. at S4386.

[35] *Id.*, 136 CONG. REC. at S4385.

[36] *Id.* § 206(b), 136 CONG. REC. at S4390.

[37] *Id.* § 206(a), 136 CONG. REC. at S4387.

[38] *Id.* § 217, 136 CONG. REC. at S4395. *See also id.* § 206(b), 136 CONG. REC. at S4392.

(twenty or more vehicles under one refueling source) in "severe" and "extreme" areas will be held to emission standards on certain pollutants that "yield when the vehicle is operated exclusively on clean alternative fuels."[39]

C. *Title III: Air Toxics*

Title III lists 191 substances and compounds that Congress declares to be Hazardous Air Pollutants and requires the Administrator to establish Minimum Emission Rates ("MERs") for each pollutant.[40] MERs vary according to the source's size.[41] Generally, these MERs may not exceed ten tons a year for any single pollutant, or twenty-five tons a year for any combination of pollutants.[42] In addition, for certain pollutants such as dioxin and mercury, the Administrator must set MERs low enough so that at least ninety percent of the sources presently exceed them.[43]

When any source exceeds its MER for a given pollutant it must implement a program designed by the Administrator to achieve maximum emissions reductions, taking into consideration cost, non-air quality related health and environmental impacts, and energy requirements.[44] The bill sets a benchmark of a ninety-percent reduction in emissions of these Hazardous Air Pollutants from uncontrolled levels.[45] Another goal of Title III is to reduce cancer caused by all air pollutants by seventy-five percent.[46]

Within three years of promulgating these first-round emissions reduction programs, the Administrator must determine if the remaining emissions threaten public health or the environment.[47] If so, the Administrator must revise the programs. Where a revised program is necessary, it must be designed without regard to cost, technological feasibility, or other economic factors, to

---

[39] *Id.* § 206(b), 136 CONG. REC. at S4388. These standards will be phased in to apply to 90% of government fleet vehicles by model year 1999, and to 30% of private fleet vehicles by model year 1997. *Id.*, 136 CONG. REC. at S4388. Credits are also available for private fleet operators who exceed the requirements. *Id.*, 136 CONG. REC. at S4389.

[40] *Id.* § 301, 136 CONG. REC. at S4399.

[41] *Id.*, 136 CONG. REC. at S4399–4400.

[42] *Id.*, 136 CONG. REC. at S4400.

[43] *Id.*

[44] *Id.*, 136 CONG. REC. at S4400.

[45] *Id.*

[46] *Id.*, 136 CONG. REC. at S4406.

[47] *Id.*, 136 CONG. REC. at S4401.

eliminate all lifetime cancer risks greater than 1-in-10,000 to "the individual in the population most exposed to emissions of a pollutant from a source."[48]

II. Title IV: Acid Rain

To better appreciate the controversy surrounding acid rain regulation, a more detailed description of this title is necessary. The stated purpose of Title IV is to regulate the two primary sources of acid rain—sulfur dioxide and nitrous oxides—by reducing the respective emissions by ten million tons and two million tons from 1980 levels.[49] The sulfur dioxide reduction plan has two phases: Phase I emissions reductions must be achieved by 1995,[50] and Phase II reductions by 2000.[51] Phase I requires 111 targeted utility plants to comply with individualized emissions limitations.[52] A plant may emit sulfur dioxide in excess of the limitations after January 1, 1995, only if it qualifies for an extension or substitution,[53] or otherwise obtains allowances for its total emissions.[54]

The allowance system is perhaps the most innovative aspect of the Amendments. An allowance is an authorization from the Administrator to emit one ton of sulfur dioxide during or after a specified year.[55] Under Phase I, the Administrator annually

---

[48] *Id.*, 136 Cong. Rec. at S4402. The steel industry alleges that this "residual risk" standard is simply unattainable by coke ovens. *Senate Clean Air Debate Shifts to Costs*, Current Dev., Env't Rep. (BNA) 1644 (Jan. 26, 1990).

Of the many other provisions in Title III, one is worthy of mention. As a corollary to restrictions on municipal incinerators, all municipalities served by an incinerator must institute a state-approved plan to recycle at least 25% of the municipal waste. S. 1630, 101st Cong., 2d Sess. § 306(a), 136 Cong. Rec. S4414 (daily ed. Apr. 18, 1990). This is the only mandatory recycling provision under the amended Clean Air Act.

[49] S. 1630, 101st Cong., 2d Sess. § 401, 136 Cong. Rec. S4418 (daily ed. Apr. 18, 1990).

[50] *Id.*, 136 Cong. Rec. at S4421–24.

[51] *Id.*, 136 Cong. Rec. at S4424–26.

[52] *Id.*, 136 Cong. Rec. at S4422–23.

[53] A two-year extension of the Phase I deadline is available, upon petition, for owners or operators of affected units who have installed coal scrubbers that reduce emissions by 90% or who have transferred their Phase I obligations to units employing scrubbers. *Id.*, 136 Cong. Rec. at S4419, S4422. A unit qualifying for an extension will receive allowances to cover the difference between the unit's actual emissions and its pre-extension limits. *Id.*, 136 Cong. Rec. at S4419, S4422.

An owner or operator may substitute its Phase I reductions by assigning them to any other unit under its control, with the permission of the Administrator. *Id.*, 136 Cong. Rec. at S4421–22.

[54] *Id.*, 136 Cong. Rec. at S4421.

[55] *Id.*, 136 Cong. Rec. at S4418.

allocates to all affected plants a specified number of allowances.[56] In addition to those allowances, excess allowances are available under a host of other provisions, the most important of which are the substitution and extension provisions.[57] The excess or bonus allowances are allocated from a reserve, which the Administrator establishes, equal to the total tonnage reduction of sulfur dioxide (not to exceed 3.5 million tons) by 1995 from all units' compliance with Phase I restrictions.[58] The Administrator may not grant substitution or extension allowances exceeding this bonus reserve. To promote the most cost-effective reduction of sulfur dioxide emissions, allowances may be transferred among lawful allowance holders, subject to EPA regulations.[59]

Phase II regulates almost all electric power generating units, with some restrictions enforceable beginning January 1, 2000.[60] In general, units which currently serve a generator with a capacity of seventy-five megawatts ("MWe") or more and which had an actual 1985 sulfur dioxide emissions rate of at least 1.2 lbs./mmBtu (pounds of sulfur dioxide per million British thermal units) are restricted from emitting more tons of sulfur dioxide than the result of the following calculation: 1.2 lbs./mmBtu multiplied by the annual average quantity of mmBtu's consumed between the years 1985 and 87, divided by 2000.[61]

There are exceptions to this basic Phase II limitation. They are designed to compensate for two undesirable results: restrictions on growth and unfair treatment of units that are already

---

[56] *Id.*, 136 CONG. REC. at S4419.

[57] *See supra* note 53. Units installed with scrubbers not only receive more time to comply, but also bonus allowances if they clean up early. If such a unit reduces its emissions below the Phase II restriction levels during the period from 1997 to 1999, it receives allowances in the amount of the excess reductions. These "double" allowances can only be described as tradeable or bankable compensation, primarily benefitting Midwestern plants, which must drastically reduce emissions by installing scrubbers. S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4421 (daily ed. Apr. 18, 1990); *see infra* text accompanying notes 89 & 102–104.

The most significant amendment to the Senate version by the conference committee is the pro rata allocation of 200,000 allowances annually in Phase I to certain utilities in Ohio, Indiana, and Illinois. In Phase II, this allocation is reduced to 50,000 annually. S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. H13,160, H13,163 (daily ed. Oct. 26, 1990).

[58] S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4421 (daily ed. Apr. 18, 1990).

[59] Id., 136 CONG. REC. at S4421.

[60] Id., 136 CONG. REC. at S4424.

[61] *Id.* This formula permits utilities to alter both the activity level and rate of emissions to find the most efficient means of achieving the Amendment's objective—fewer tons of sulfur dioxide emissions. *See* S. REP. No. 228, 101st Cong., 1st Sess. 322 (1989).

clean. Because the main Phase II emission standards are based on 1985 and 1987 use levels, they could unfairly limit emissions and economic growth if 1985 to 1987 was an unusually low use period in certain regions. Some regions could be permanently restricted to a recession level energy supply.[62] A compromise measure allows units which operated at less than sixty percent of capacity in 1985 to receive bonus allowances after the year 2000.[63]

In contrast to its relatively generous treatment of existing plants, Phase II is less accommodating for growth of new power plants. For example, most plants that begin operation after 1990 will receive allowances under the Phase II formula using an emission rate of only 0.3 lbs./mmBtu.[64] More importantly, they will receive allowances to operate at only sixty-five percent of capacity.[65] This requires the units to purchase thirty-five percent of their capacity allowances from other holders, operate a superclean facility, or operate at less than capacity. Units which do not begin operation until after enactment of the Amendments must purchase *all* allowances from the other holders.

The Phase II scheme allows some growth for plants that are already clean. Plants with an emission rate below 1.2 lbs./mmBtu in 1985 may remain at that level until January 1, 2000.[66] After that date, the Administrator can issue them an additional twenty percent of their annual allowances.[67] The "Clean States" provision allocates 125,000 allowances annually among units within all states with an average emission rate of 0.8 lbs./mmBtu or less.[68] Clean states are thereby assured of some excess allowances to accommodate growth, even if there is a shortage of allowances for sale in the market.

The Amendments further promote allowance trading to accommodate growth,[69] by providing EPA-conducted sales and auctions of allowances. The Administrator must make available for sale, at $1,500 each (adjusted for inflation), 100,000 one-ton

---

[62] *Acid Rain: Hearings on S. 1630 Before the Subcomm. on Environmental Protection of the Senate Comm. on the Environment and Public Works*, 101st Cong., 1st Sess., pt. 5, 61 (1989) (statement of Senator Dan Coats (R-Ind.)) [hereinafter *Hearings*].

[63] S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4424 (daily ed. Apr. 18, 1990).

[64] *Id.*, 136 CONG. REC. at S4426.

[65] *Id.*, 136 CONG. REC. at S4425–26.

[66] *Id.*, 136 CONG. REC. at S4425.

[67] *Id.*

[68] *Id.*, 136 CONG. REC. at S4426.

[69] *See infra* text accompanying notes 104–106.

allowances usable for each calendar year beginning in 2000.[70] Beginning in 1995, the Administrator must also conduct semi-annual auctions of 100,000 allowances.[71] The 100,000 allowances may include allowances unsold after the fixed-price sale, and may be supplemented by allowances contributed for sale by private holders.[72]

As in Phase I, all allowances in excess of the initial emissions limitations come from a reserve fund. The Administrator must establish a Phase II allowance reserve equal to the total sulfur dioxide emissions reduction achieved in year 2000, from compliance with Phase II restrictions, not to exceed 5.3 million allowances.[73] The Administrator is to distribute ten percent of these allowances annually from 2000 to 2009,[74] according to the Phase II special provisions discussed above.[75] If the reserve is not large enough to accommodate all the allowances earned under other provisions, the Administrator must deduct from each plant its pro rata share of the difference. The Administrator may not deduct allowances from the auction and sales programs. After January 1, 2000, the Administrator may not allocate allowances exceeding the annual emissions cap of 8.9 million tons.[76]

The teeth of the acid rain title is a penalty of $2,000 per ton of emission exceeding a given year's allowances, and a requirement that plants offset excess sulfur dioxide emissions the following year.[77] This is a crucial part of Title IV because noncompliance must be prohibitively unattractive if the Amendments are to achieve their objectives.[78]

---

[70] S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4419 (daily ed. Apr. 18, 1990).

[71] *Id.*, 136 CONG. REC. at S4420. The conference committee reduced the number of allowances offered for sale to 50,000 annually and increased the number at auction to 150,000 from 1993 to 1995, and 250,000 after 1995. S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. H13,170–71 (daily ed. Oct. 26, 1990).

[72] S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4419 (daily ed. Apr. 18, 1990).

[73] *Id.*, 136 CONG. REC. at S4424.

[74] *Id.*

[75] *See supra* notes 63–72 and accompanying text.

[76] S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4419 (daily ed. Apr. 18, 1990).

[77] *Id.*, 136 CONG. REC. at S4429.

[78] *See Hearings*, *supra* note 62, pt. 5, at 196 (statement of Daniel J. Dudek, Senior Economist, Environmental Defense Fund).

Title IV also restricts nitrous oxide emissions. One set of emission rates must be met by any unit affected by the sulfur dioxide requirements. S. 1630, 101st Cong., 2d Sess. § 401, 136 CONG. REC. S4426–27. The Administrator must promulgate a second round of limitations effective in 1997 which will be based on available technology, energy and environmental impacts, as well as costs comparable to first-round controls. *Id.*, 136 CONG. REC. at S4429.

III. The Sulfur Dioxide Emissions Debate

The debate over the acid rain provisions, like the debate over much environmental legislation, revolved around two questions: (1) is the environmental damage significant enough to justify the cost of eliminating it, and (2) even if the cost is justified, how should that cost be spread around the country?

Regarding the first question, practically everyone agrees that some action should be taken to control acid rain,[79] but there is no consensus on the amount of damage that acid rain causes. Preliminary results of the decade-long National Acid Precipitation Assessment Program ("NAPAP") state that precipitation is the dominant cause of acidity in seventy-five percent of the acidified lakes and fifty percent of the acidified streams in this country.[80] Five percent of Northeastern lakes are acidic, including eleven percent in the Adirondacks.[81] In the Northeast, a thirty-percent reduction in sulfur dioxide emissions would yield a fifty-percent recovery of acidified lakes in fifty years.[82] Continued sulfur dioxide emissions at current levels would turn twenty percent of the lakes in the Blue Ridge Mountain region acidic, but would not change the number of acidified lakes in the Northeast.[83]

These results are not conclusive, however. NAPAP Director James R. Mahoney conceded that predicting effects on lakes from changes in emissions is not a precise science.[84] Many scientists in the field believe the NAPAP conclusions understate the acid rain problem.[85] One study attributes $5 billion in crop loss annually to acid rain.[86] At the other extreme, some who have described acidic lakes in Australia as "fishless but highly prized," with "no green slime or leeches," simply are not con-

---

[79] Even the National Coal Association, which expects to suffer the most as utilities switch from coal to natural gas, is resigned to some acid rain legislation. *See Hearings*, *supra* note 62, pt. 5, at 31 (statement of Richard Lawson, President of National Coal Association).

[80] *Acid Rain's Role in Lake, Stream Acidity, Other Effects Documented in First NAPAP Draft*, Current Dev., Env't Rep. (BNA) 844 (Aug. 31, 1990).

[81] *Hearings*, *supra* note 62, pt. 6, at 6–7 (testimony of James R. Mahoney, Director of NAPAP). *See also* N.Y. Times, Dec. 31, 1989, at 1, col. 3.

[82] *Hearings*, *supra* note 62, pt. 6, at 7.

[83] *Id.* at 7, 10.

[84] *Id.* The 95% confidence interval on the prediction that 161 lakes would remain acidic at the current level of emissions was ±245 lakes. *Id.* at 42 (Table 4).

[85] *See* N.Y. Times, Mar. 20, 1990, at C4, col. 1.

[86] *Ozone, Acid Rain Cause Extensive Damage to U.S. Crops, Forests, WRI Says in Report*, Current Dev., Env't Rep. (BNA) 1779 (Feb. 16, 1990).

vinced that acid rain is a harmful phenomenon.[87] The analysis that follows, however, assumes some agreement on the harm of acid rain and concentrates instead on the second question, how to distribute the costs.

A. *The Dominance of Regional Interests*

Regional interests, much more than party politics, dominate the acid rain debate. It has been said, "[t]ell me where you stand on acid rain, and I'll know where you live within fifty miles."[88] Most importantly, the interests of the Midwest and Appalachia oppose those of the rest of the country in two ways. First, these areas have the combined characteristics of few clean plants and a tendency to burn high-sulfur coal. High emissions[89] mean that this area will incur the greatest costs to bring emissions in line with the national Phase I and II standards.

Second, this area mines most of the high-sulfur coal that it uses.[90] Some plants in the area will decide to reach the emissions limits by using scrubbers, but others will decide to burn low-sulfur coal instead.[91] One estimate predicts that a switch to low-sulfur coal will result in a loss of 22,000 to 30,000 mining and support jobs with a consequent $3.5 billion loss of direct and indirect income.[92] The situation becomes more bleak considering that displaced miners in Pennsylvania, Ohio, and West Virginia have no clout to urge other states that purchase their high-sulfur coal to use scrubbers instead of switching to low-sulfur coal.

For these reasons, senators, particularly Robert Byrd (D-W.Va.), strongly urged a cost-sharing provision.[93] One proposal was for a nationwide tax on sulfur dioxide emissions, the proceeds to be distributed to dirty utilities to help offset the cost

---

[87] *Conservative Coalition Criticizes Air Bill, Says No Harmful Effects Come from Acid Rain*, Current Dev., Env't Rep. (BNA) 2002 (Apr. 27, 1990). *Cf. The Simpsons* (FOX television broadcast, Nov. 1, 1990) (Gubernatorial candidate and nuclear power plant president Montgomery Burns describes a three-eyed fish caught near the reactor as a genetically advanced superfish).

[88] *Offsetting the Cost of Clean Air*, PUB. UTIL. FORT., Apr. 12, 1990, at 8 (quoting unidentified state legislator).

[89] *See* S. REP. No. 228, 101st Cong., 1st Sess. 284 (1989) (Figure IV-10).

[90] *See* S. 1630, 101st Cong., 1st Sess., 135 CONG. REC. S1767, (daily ed. Feb. 28, 1989) (remarks of Senator Dave Durenberger (R-Minn.).

[91] *Hearings*, *supra* note 62, pt. 5, at 38 (remarks of David Hawkins, Senior Staff Attorney, National Clean Air Coalition).

[92] *Id.* at 25 (statement of Richard Trumka, President, United Mine Workers of America). *See also* N.Y. Times, Apr. 1, 1990, at A20, col. 3.

[93] *See infra* note 99.

of technology necessary to meet the emissions restrictions.[94] Since cost-sharing exclusively benefits states with high sulfur dioxide emissions, states concentrated in the Midwest and Appalachia,[95] the proposals met staunch opposition from states that believed a state should not benefit because it is a high polluter.[96]

In addition to opposing cost-sharing, the West has its own particular agenda: it wants to assure that the emissions allowance system will accommodate its expected economic and population growth. In the past, each one-percent increase in Gross National Product ("GNP") has resulted in approximately a one-percent increase in demand for electricity.[97] Westerners fear that an emissions cap and transferability barriers will not permit the flow of allowances necessary to sustain Western growth.

The East, especially the Northeast, has the dubious distinction of being the primary recipient of acid rain produced in the United States.[98] Thus, Easterners are the strongest supporters of a strict acid rain provision, and are likely not to be sympathetic to cost-sharing proposals to aid the Ohio Valley plants that have been supplying the Adirondacks with acid rain for years.

B. *Compromises Among Regions Through the Allowance Program*

The challenge for the drafters of the Amendments was accommodating all of these regional interests in a single statute. The most significant hurdle to overcome was to accommodate the Midwestern needs to ease the costs of the Amendments without resort to the politically unpalatable emissions tax or direct aid to laid-off miners.[99] The Senate directly stated in S. 1630 its

---

[94] *See, e.g., Hearings, supra* note 62, pt. 5, at 192 (remarks of Senator Paul Simon (D-Ill.)).

[95] *Id.* at 222 (chart used in remarks of Senator Joseph Lieberman (D-Conn.)).

[96] See N.Y. Times, Jan. 11, 1990, at A19, col. 1 (governors or other officials from Arizona, Colorado, Massachusetts, Minnesota, Montana, New Hampshire, New Jersey, North Carolina, Utah, Wisconsin, and Wyoming form coalition opposing cost-sharing).

[97] *Hearings, supra* note 62, pt. 5, at 106 (written statement of Richard Lawson, President, National Coal Association).

[98] *Id.*, pt. 6, at 38 (written statement of Mr. Mahoney, Table 1).

[99] *See, e.g., Byrd Amendment to Aid Unemployed Coal Miners Falls Short in 50-49 Senate Clean Air Bill Vote*, Current Dev., Env't Rep. (BNA) 1939 (Apr. 6, 1990) (Senator Byrd's amendment giving three-years' compensation to miners of high-sulfur coal displaced by the bill narrowly rejected).

distaste for cost-sharing programs.[100] The compromise is the provision granting allowances to the extent that an eligible plant outperformed the Phase I requirements.[101] Since Midwestern plants will achieve the most drastic reductions in emissions, this provision primarily benefits them by partially rewarding them for any emissions-reducing technology they install. The allowances may be banked to accommodate future energy growth, or they may be sold. Either way, the provision acts like the direct aid that Midwestern senators requested.[102]

By essentially manufacturing allowances to distribute to Midwestern plants, Congress also intended to accommodate the growth interests of the West. While not selling the Midwest off as economically stagnant, the Amendments must presume that the Midwest will not need all the allowances that it can easily obtain by reducing its emissions.[103] It will be attractive to sell these allowances to the West to help offset the cost of control technologies.[104]

Many Westerners were not satisfied with this answer to their concerns. As one member of Congress asked, "[d]o we reward the last in the nation to clean up by allowing them to decide who gets to grow in the future?"[105] In response, the Senate added the fixed price sale and auction provisions, thus guaranteeing a minimum availability of allowances.[106]

C. *The Effectiveness of These Compromises in Promoting Allowance Trading*

The primary concerns of the growth states are transferability problems from imperfect information, and hoarding of allow-

---

[100] S. 1630, 101st Cong., 2d Sess. § 407, 136 CONG. REC. S4431 (1990).

One senator astutely noted that taxpayers across the country share in the bailout of failed savings and loans located primarily in the Southwest, suggesting cost-sharing is not such a selfish request. *Hearings*, *supra* note 62, pt. 5, at 60 (statement of Senator Coats).

[101] *See supra* note 57.

[102] *See Energy Policy Implications of the Clean Air Act of 1989: Hearings Before Senate Comm. on Energy and Natural Resources*, 101st Cong., 2d Sess. 28, 55 (1990) (statement of Linda G. Stuntz, U.S. Department of Energy).

[103] *See generally Hearings*, *supra* note 62, (remarks of Mr. Dudek).

[104] *Cf. id.*, pt. 5, at 204 (remarks of William A. Badger, Chairman, Commission on Electricity, National Association of Regulatory Utility Commissioners).

[105] *House, Senate Head in Different Directions over Air Bill Acid Rain Cost-Sharing Provision*, Current Dev., Env't Rep. (BNA) 1772, 1774 (Feb. 16, 1990) (comments of Representative W.J. Tauzin (D-La.)). *See also Hearings*, *supra* note 62, pt. 5, at 226 (statement of Colorado Governor Roy Romer on behalf of the Alliance for Acid Rain Control).

[106] *See supra* notes 69–72 and accompanying text.

ances by Midwestern states to protect their own growth opportunities. By putting the EPA in the position of allowance broker, the auction and sales provisions will help overcome obstacles of transaction costs and imperfect market communications. Also, by keeping participation in the auction partially voluntary, the program will still permit an arrangement such as joint investments on scrubber installation, in which the clean utility assigns a portion of its allowances to the utility which contributed the capital for the scrubber installation.[107]

The auction and sales provisions may not completely cure the hoarding problem, however. Assuming that the utilities have perfect information, the basic requirement for an allowance exchange is that the cost to the seller (the sum of the marginal cost of cleaning to obtain an allowance and the growth opportunity cost of not emitting that ton of pollution in the future) must be less than the price the buyer is willing to pay for the allowance (the marginal value of the growth available to the purchaser from use of the allowance). The idea behind the Amendments is that growth in the Midwest will not be so great as to drive the marginal growth opportunity cost to a level where it is an impediment to selling bonus allowances. With that factor out of the equation, as long as the marginal cost of cleaning exceeds the purchaser's marginal value of growth, there will be allowance trading.[108]

Two important questions arise from the allowance market that the Amendments create. First, what may occur if, as Western senators implicitly allege,[109] state commissioners overestimate their growth opportunity costs and bank their allowances instead of selling them? Second, what effect will the addition of auction and sales of Phase II allowances have on the model?

The worry, stated often in the hearings on S. 1630,[110] is that state commissioners will forbid utility operators from selling allowances to other states even though the offer price of the allowance exceeds that state's marginal growth opportunity cost. This is quite conceivable given certain political factors. No state governor believes or would want to admit that his

---

[107] *See Hearings*, *supra* note 62, pt. 5, at 207 (remarks of Mr. Dudek).

[108] Congress heard no testimony on future growth in the Midwest; the validity of any empirical assumption about Midwestern growth is beyond the scope of this Essay.

[109] *See supra* note 105.

[110] *See Hearings*, *supra* note 62, pt. 5, at 198, 211, 213 (statements of Mr. Badger and William Walbridge, Executive Vice-President, Seminole Electric Co-operative).

state's economy will not grow; state agencies' estimates of growth likely will exceed actual growth. This "puffing" of the state's economy will retard the sales of some allowances that otherwise would be sold in the model.

Another fear is that state politicians will not only miscalculate their marginal growth opportunity cost, but will consciously inflate it because of the potentially career-ending repercussion of an under-estimate. Imagine the following scenario. All available data suggests that state *A* has a marginal growth opportunity cost that exceeds the market price for allowances for its first 10,000 annual allowances from 1995 to 2000. If it will receive 12,000 allowances annually according to the Amendments, it should sell 2000 allowances annually and use the proceeds to lower its residents' energy rates. However, the estimate of 10,000 is not precise. There may, for example, be a twenty-percent chance that the level at which price exceeds cost will be either 7500 or 12,500 allowances, and a five-percent chance the level will be either 6000 or 14,000. If some unpredictable boom in the economy drives demand for energy in state *A* above 10,000, the politician who sold state *A*'s allowances is doomed because state *A* may be forced to purchase new allowances after the market price has risen. Nothing could be more damaging politically than having sold your state's opportunity for growth to some other state. The Governor may inflate the need for allowances to 14,000 to avoid any risk of personal political disaster.

The allowance provision, on the other hand, is somewhat obscure to the press and electorate. A failure to sell all the available allowances will not meet significant political backlash. Indeed, the electorate, wearing the same rose-colored glasses as the Governor, may even agree that the state should retain the excess allowances.

The combined factors of puffing and political overcompensation will certainly have an effect on the perceived marginal growth opportunity cost. Perhaps the only way to eliminate these effects would be to tighten the duty that utility operators owe to ratepayers to keep energy costs as low as possible.[111] A means of strengthening such a duty would be to permit a citi-

[111] *See Hearings*, *supra* note 62, pt. 5, at 210 (testimony of Mr. Dudek that if states force utilities to "restrict trading, the utilities will only drive up costs to their ratepayers").

zens' suit to enjoin a governor or appointed state power commissioner from banking allowances.

The auction and sales provisions partially solve the hoarding problem. They ensure that at least 200,000 allowances will be available each year from 1993 to 1995, and that 300,000 allowances will be available each year after 1995. To that extent, at least, growing states will be able to purchase needed allowances.[112]

The auction and sales provisions may also have an indirect effect on Phase I allowance transfers. The auction program will provide important price information that will enlighten open market traders as to the future value of allowances. Of course, if allowances are being freely traded in Phase I, the market price will determine the auction price, not vice-versa. Auction price will be the discounted value of the current open market price. But if Phase I trading is not occurring, the first auction in 1995 will yield a futures price of an allowance usable in 2000 from which an efficient price could be derived for allowances currently usable. In this way, information barriers to Phase I trading may be overcome by the auctions.

Using the market price that emerges from the 1995 auctions, utilities will also be able to determine more easily the merit of further reducing emissions to obtain more Phase II allowances. After 1995, utilities will not get double allowances for emissions reductions, but they still may have saleable allowances by emitting less than their limitation. Under Phase II, at least 100,000 allowances must be offered for sale annually from the reserve regardless of utilities' desire to bank the allowances.[113] With a market price for year-2000 allowances assured,[114] utilities can confidently decide if their marginal cost of reducing emissions by one ton exceeds the auction price paid for an allowance

---

[112] The April 3, 1990, Senate version of the Amendments had a serious flaw: no allowance transferred under either the auction or sales program was usable until after the year 2000. Under the Senate version, a Phase I allowance sold on the open market could have been used in or after the designated year. *See supra* note 70. Accordingly, a Phase I allowance holder would not have utilized the auction before 2000 because Phase I allowances sold on the market would have had greater liquidity (and accordingly greater value) than those sold at auction and then banked until 2000. Phase I allowances would have been sold on the open market where a higher price would have captured their liquidity.

[113] *See supra* text accompanying notes 66–71.

[114] Conceivably, without the auction risk-averse allowance holders could overestimate their marginal growth opportunity costs, hoard allowances, and prevent establishment of a market price.

usable in 2000. Admittedly, they still must determine that the auction price exceeds their state's marginal growth opportunity cost. But it is more likely that an accurate calculation will prevail over political rhetoric if true market prices are available for comparison.

IV. CONCLUSION

The Clean Air Act Amendments of 1990 are the first significant attempt at strengthening clean air legislation in over a decade. Despite several attempts in Congress during this period to initiate reform legislation, all efforts failed. The 101st Congress has overcome the stalemate because rising environmental consciousness has weakened industrial and regional lobbyists and has urged a Republican President to support clean air legislation. Nevertheless, the battles rage as to how strict the legislation should be.

In the battle over acid rain, competing interests constructed a market for sulfur dioxide emission allowances that could accommodate disparate growth interests and clean-up costs within a framework of bottom-line emissions reductions. Ideally, this constructed market guarantees that the reductions are achieved at the lowest cost. Congress inserted the auction and sales provisions to ensure that utilities freely trade allowances. This brief examination of a few scenarios demonstrates that provisions in the Amendments do encourage an active allowance market, and may even provide some guarantees to market information, but unforeseen political and macroeconomic factors could nevertheless have damaging effects on the allowance program.

—Brian L. Ferrall

THE ROLE OF STATE COURTS IN NARROWING OVERBROAD SPEECH LAWS AFTER *OSBORNE V. OHIO*

> When speech is eloquent and the ideas lofty, it is easy to find restrictions on them invalid. But were the First Amendment limited to such discourse, our freedom would be sterile indeed. Mr. Osborne's pictures may be distasteful, but the Constitution guarantees both his right to possess them privately and his right to avoid punishment under an overbroad law.[1]

Recent controversies over flag burning,[2] obscene song lyrics,[3] funding for the arts,[4] and pornography[5] have spurred both federal and state governments to consider and enact legislation designed to curb free speech. In the past, similarly controversial speech laws have been successfully challenged as overbroad and vague.[6] While overbreadth and vagueness often "go hand

---

[1] *Osborne v. Ohio*, 110 S. Ct. 1691, 1717 (1990), *reh'g denied*, 110 S. Ct. 2605 (1990) (Brennan, J., dissenting).

[2] United States v. Eichman, 110 S. Ct. 2404 (1990); Texas v. Johnson, 109 S. Ct. 2533 (1989).

[3] Louisiana Governor Buddy Roemer vetoed a bill that would have required warning labels on recordings that "promote" deviant sex, violence, drug abuse, suicide, or child abuse. Washington Post, July 26, 1990, at C1, col. 1. In Florida, the rap group 2 Live Crew was acquitted of obscenity charges stemming from a live performance of songs from their album "As Nasty as They Wanna Be." *In Rap Music, the Beat and the Lawsuits Go On*, N.Y. Times, Oct. 23, 1990, at C13, col. 1. A Fort Lauderdale, Florida, record store owner, however, was convicted of selling the album to an adult customer. *Id.* at C13. *See* Holt, *Protecting America's Youth: Can Rock Music Lyrics Be Constitutionally Regulated?* 16 J. CONTEMP. LAW 53 (1990); Comment, *Regulating Rock Lyrics: A New Wave of Censorship?* 23 HARV. J. ON LEGIS. 595 (1986).

[4] After much heated debate, Congress recently passed legislation extending the life of the National Endowment for the Arts ("NEA") for three more years. The legislation allows the NEA in reviewing arts grants to consider "general standards of decency and respect for the diverse beliefs and values of the American public." *The Endowment's New Lease on Life*, N.Y. Times, Nov. 3, 1990, at 24, col. 1. *See Cincinnati Jury Acquits Museum in Mapplethorpe Obscenity Case*, N.Y. Times, Oct. 6, 1990, at 1, col. 1.

[5] In American Information Enters. v. Thornburgh, 742 F. Supp. 1255 (S.D.N.Y. 1990), a federal district judge barred enforcement of a "dial-a-porn" law sponsored by Senator Jesse Helms (R-N.C.), because its use of the term "indecent" was vague and overbroad. The "Helms Amendment" would have prevented telephone services that provide indecent, but not obscene, messages to persons under 18, and would have required adults to subscribe to the services in writing. *See* Note, *The Congressional Response to the Supreme Court's Treatment of Dial-A-Porn*, 78 GEO. L.J. 2025 (1990); Note, *Telephone Pornography: First Amendment Constraints on Shielding Children from Dial-A-Porn*, 22 HARV. J. ON LEGIS. 503 (1985).

[6] *See* Board of Airport Comm'rs of Los Angeles v. Jews for Jesus, Inc., 482 U.S. 569 (1987) (invalidating rule that prohibited all first amendment activities in Los Angeles airport); Houston v. Hill, 482 U.S. 451 (1987) (invalidating ordinance against interrupting a police officer in the execution of his duty).

in hand,"[7] this Essay will focus on overbroad laws. The public debate over the enactment and challenge of overbroad laws is often vociferous, yet little attention is given to the process by which courts and legislatures attempt to "save" these laws by narrowing their applicability when they are challenged.

While the Supreme Court has restricted the ability of federal courts to narrow overbroad speech laws enacted by Congress, the Court seems to have opened the door for state courts to engage in substantial narrowing, and indeed rewriting, of speech laws enacted by their respective state legislatures. This Essay examines how the Supreme Court has opened this door in *Osborne v. Ohio*,[8] and the potential problems of giving state courts such broad quasi-legislative power. It concludes that unless clear legislative intent can be discerned, state courts should invalidate overbroad speech laws rather than attempt to narrow them.

I. THE *OSBORNE* AND *OAKES* DECISIONS

First amendment overbreadth doctrine is premised on the idea that laws which have the potential to chill the protected expression of individuals not before the court should be void.[9] As such, overbreadth doctrine is a traditional exception to the notion that "individuals may not litigate the rights of third parties."[10] Like other areas of constitutional adjudication, cases involving "distasteful" activities,[11] most notably child pornography, have established the parameters of individual rights for ordinary people.

Hidden in the Supreme Court's landmark decision, *Osborne v. Ohio*—a decision which allowed states to penalize the mere

---

[7] *See* Local 189 Int'l Union of Police Ass'ns v. Barrett, 524 F. Supp. 760, 765 (N.D. Ga. 1981).

[8] 110 S. Ct. 1691 (1990).

[9] Thornhill v. Alabama, 310 U.S. 88, 97 (1940). Under Broadrick v. Oklahoma, 413 U.S. 601, 615 (1973), the overbreadth must be "substantial" for a court to strike down the law. For a comprehensive treatment of overbreadth doctrine, see Monaghan, *Overbreadth*, 1981 S. CT. REV. 1; Note, *The First Amendment Overbreadth Doctrine*, 83 HARV. L. REV. 844 (1970).

[10] L. TRIBE, AMERICAN CONSTITUTIONAL LAW 1023 (2d ed. 1988). *See* NAACP v. Button, 371 U.S. 415, 432 (1963) ("the instant decree may be invalid if it prohibits privileged exercises of First Amendment rights whether or not the record discloses that the petitioner has engaged in privileged conduct"); Naturist Soc'y, Inc. v. Fillyaw, 736 F. Supp. 1103, 1110 (S.D. Fla. 1990) ("the traditional rules of standing are relaxed to prevent an overbroad law from becoming a de facto prior restraint on speech").

[11] *Osborne*, 110 S. Ct. at 1717 (Brennan, J., dissenting).

possession and viewing of child pornography[12]—is an implicit invitation to state courts to engage in substantial narrowing of overbroad speech laws. Though five Justices had ruled one year earlier in *Massachusetts v. Oakes*[13] that state legislatures could not cure their overbroad laws by a saving construction or subsequent amendment, the Supreme Court inexplicably affirmed the power of state courts to do essentially the same thing by allowing them to narrow a law after it is challenged.

In *Oakes*, the respondent had taken ten color photographs of his fourteen-year-old stepdaughter "sitting, lying, and reclining on top of a bar, clad only in a red and white striped bikini panty and a red scarf."[14] The breasts of the stepdaughter—who was described by the Court as "physically mature"[15]—were fully exposed in the photographs. Douglas Oakes was convicted under a statute which prohibited anyone from allowing a minor "to pose or be exhibited in a state of nudity."[16] The Massachusetts Supreme Judicial Court reversed Oakes' conviction. After the U.S. Supreme Court granted certiorari, the Massachusetts legislature, hoping to eliminate Oakes' overbreadth challenge, amended the statute to require "lascivious intent." Relying on *Bigelow v. Virginia*,[17] Justice O'Connor's plurality opinion found that the intervening legislative amendment rendered Oakes' overbreadth challenge moot. But in a separate opinion Justice Scalia, writing for himself and four other Justices, argued that such a subsequent amendment could not eliminate an overbreadth defense:

---

[12] *Osborne* substantially limited the holding of Stanley v. Georgia, 394 U.S. 557 (1969). At the time *Osborne* was decided, 19 states had prohibited the possession of child pornography, though the federal government has never enacted such a statute. *Osborne*, 110 S. Ct. at 1713–14 n.17 (Brennan, J., dissenting).

[13] 109 S. Ct. 2633 (1989).

[14] *Id.* at 2636.

[15] *Id.*

[16] MASS. GEN. L. ch. 272, § 29A (1986). Another part of the statute defined nudity as:

> uncovered or less than opaquely covered human genitals, pubic areas, the post-pubertal human female breast below a point immediately above the top of the areola, or the covered male genitals in a discernibly turgid state . . . . In the case of pre-pubertal persons nudity shall mean uncovered or less than opaquely covered pre-pubertal human genitals or pubic area.

Id. § 31.

[17] 421 U.S. 809 (1975). In *Bigelow*, a statute which prohibited encouraging abortions by "publication, lecture [or] advertisement" was challenged as overbroad. *Id.* at 812–13. Because the law was amended after the defendant's conviction, the Court concluded that the overbreadth issue was moot, since the original law could no longer chill protected speech. *Id.* at 818.

> It seems to me strange judicial theory that a conviction initially invalid can be resuscitated by postconviction alteration of the statute under which it was obtained. Indeed, I would even think it strange judicial theory that an act which is lawful when committed (because the statute that proscribes it is overbroad) can become retroactively unlawful if the statute is amended *preindictment*.[18]

The plurality, however, remanded the case to determine whether the former version of the statute could be applied constitutionally to Douglas Oakes.[19]

In *Osborne*, the petitioner was found with four photographs of a nude male adolescent in sexually explicit positions.[20] Clyde Osborne was convicted under an Ohio statute prohibiting the "possess[ion] or view[ing of] any material or performance that shows a minor who is not the person's child or ward in a state of nudity . . . ."[21] Because mere nudity constitutes protected first amendment expression,[22] the Ohio Supreme Court narrowly

---

[18] *Oakes*, 109 S. Ct. at 2639 (Scalia, J., concurring and dissenting) (emphasis in original). Justices Brennan, Blackmun, Marshall, and Stevens joined Justice Scalia in this part of the opinion. Only Justice Blackmun agreed with Justice Scalia that the case should be remanded to dispose of the as-applied challenge, since the law was not unconstitutionally overbroad. *Id.* at 2640–41. Justice Brennan's dissent argued that the former version of the statute was fatally overbroad. *Oakes*, 109 S. Ct. at 2646 (Brennan, J., dissenting).

[19] *Oakes*, 109 S. Ct. at 2639.

[20] It was unclear whether all four photographs were of one boy. "Three photographs depict the same boy in different positions: sitting with his legs over his head and his anus exposed; lying down with an erect penis and with an electrical object in his hand; and lying down with a plastic object which appears to be inserted in his anus." *Osborne*, 110 S. Ct. at 1695 n.1. Osborne testified that the boy was 14 at the time of the photographs. The fourth photograph depicted only the torso of a nude, standing boy. *Id.* at 1695 n.1.

[21] *Osborne*, 110 S. Ct. at 1694–95. The full statute provides that:

> (A) No person shall do any of the following:
>
> . . . .
>
> (3) Possess or view any material or performance that shows a minor who is not the person's child or ward in a state of nudity, unless one of the following applies:
>
> (a) The material or performance is sold, disseminated, displayed, possessed, controlled, brought or caused to be brought into this state, or presented for a bona fide artistic, medical, scientific, educational, religious, governmental, judicial, or *other proper purpose*, by or to a physician, psychologist, sociologist, scientist, teacher, person pursuing bona fide studies or research, librarian, clergyman, prosecutor, judge, or other person having a proper interest in the material or performance.
>
> (b) The person knows that the parents, guardian, or custodian has consented in writing to the photographing or use of the minor in a state of nudity and to the manner in which the material or performance is used or transferred.

OHIO REV. CODE ANN. § 2907.323(A)(3) (Supp. 1989) (emphasis added).

[22] New York v. Ferber, 458 U.S. 747, 765 n.18 (1982). *See* Erznoznik v. Jacksonville, 422 U.S. 205 (1975). *But see* Naturist Soc'y v. Fillyaw, 736 F. Supp. 1103, 1111 (S.D. Fla. 1990) ("public nudity alone has no First Amendment protection").

read the statute as applying only to "the possession or viewing of material or performance of a minor who is in a state of nudity, where such nudity constitutes *a lewd exhibition or involves a graphic focus on the genitals*, and where the person depicted is neither the child nor the ward of the person charged."[23] The Ohio court also found that the statute's silence about scienter was not unconstitutional.[24]

The clear purpose of the Ohio court's narrowing was to save the statute from an overbreadth claim that it punished people for possessing or viewing innocent nude photos of children.[25] The U.S. Supreme Court found the statute not impermissibly overbroad as construed by the Ohio Supreme Court, but nevertheless reversed Osborne's conviction and remanded the case, because it was not clear that the state had proven each of the elements required under the Ohio statute.[26]

While *Oakes* and *Osborne* may seem consistent at first glance—the Supreme Court rejected both overbreadth challenges, yet remanded both cases on due process grounds—a more careful reading indicates a subtle, but important, difference. In *Oakes*, five Justices recognized the need to prevent state legislatures from saving their overbroad statutes by subsequent amendment.

> The overbreadth doctrine serves to protect constitutionally legitimate speech not merely *ex post*, that is, after the offending statute is enacted, but also *ex ante*, that is, when the legislature is contemplating what sort of statute to enact. If the promulgation of overbroad laws affecting speech was cost free . . . that is, *if no conviction of constitutionally proscribable conduct would be lost, so long as the offending statute was narrowed before the final appeal—then legislatures would have significantly reduced incentive to stay within constitutional bounds in the first place.* When one takes account of those overbroad statutes that are never

---

[23] State v. Young, 37 Ohio St. 3d 249, 252, 525 N.E.2d 1363, 1368 (1988) (emphasis added).

[24] *Id.*, 37 Ohio St. 3d at 252, 525 N.E.2d at 1368. The Ohio court cited section 2901.21(B), which provides that recklessness is the appropriate standard when a statute is silent about mens rea.

[25] In his *Oakes* dissent, Justice Brennan argued that such nude photos were taken every day and noted that great artists, such as Degas, Renoir, and Donatello, painted nude children. *Oakes*, 109 S. Ct. at 2643–46 (Brennan, J., dissenting). Osborne's trial counsel even suggested that under the Ohio statute an individual "probably couldn't even have nude photographs of himself." *Osborne*, 110 S. Ct. at 1704.

[26] *Osborne*, 110 S. Ct. at 1703–05. Justice Scalia voted with the majority in *Osborne*. Justice White, who wrote the *Osborne* majority, joined Justice O'Connor's plurality in *Oakes*.

> challenged, and of the time that elapses before the ones that are challenged are amended to come within constitutional bounds, a substantial amount of legitimate speech would be "chilled" as a consequence of the rule the plurality would adopt . . . I have heard of a voidable contract, but never of a voidable law. The notion is bizarre.[27]

This same distrust of state legislatures was evident in Justice White's majority opinion in *Osborne*: "[l]egislatures who know they can cure their own mistakes by amendment without significant cost may not be as careful to avoid drafting overbroad statutes as they might otherwise be."[28] In its haste to reaffirm its distrust of state legislatures, however, the Court mistakenly placed its trust in state courts to remedy overbroad laws:

> But a similar effect will not be likely if a judicial construction of a statute to eliminate overbreadth is allowed to be applied in the case before the Court. This is so primarily because the legislatures cannot be sure that the statute, when examined by a court, will be saved by a narrowing construction rather than invalidated for overbreadth. In the latter event, there could be no convictions under that law even of those whose own conduct is unprotected by the First Amendment. Even if construed to obviate overbreadth, applying the statute to pending cases might be barred by the Due Process Clause. Thus, careless drafting cannot be considered to be cost free based on the power of the courts to eliminate overbreadth by statutory construction.[29]

Whether judicial construction solves the problem of overbroad laws or creates its own problems is the subject of Part II.

II. Analysis

While the power of courts to narrow overbroad statutes has long been recognized,[30] this power is not without limitation. In the federal context, the Supreme Court has been "mindful that the lawmaking power lies with Congress, and that there is a difference between adopting a saving construction and rewriting

---

[27] *Oakes*, 109 S. Ct. at 2639–40 (Scalia, J., concurring and dissenting) (emphasis added).

[28] *Osborne*, 110 S. Ct. at 1702.

[29] *Id.*

[30] *See* Erznoznik v. Jacksonville, 422 U.S. 205, 216 (1975) ("the Court has held that a[n overbroad] state statute should not be deemed facially invalid unless it is not readily subject to a narrowing construction by the state courts").

legislation altogether."[31] In *United States v. Reese*,[32] the Court noted that "introduc[ing] words of limitation" into an overbroad criminal statute would "substitute the judicial for the legislative department," in effect creating a new law, not enforcing an old one.[33] In *Scales v. United States*,[34] the Court cautioned against "perverting the [legislative] purpose of a statute" in attempting to narrow overbroad laws.[35]

In the state context, however, the Supreme Court has provided less guidance to the state courts on their appropriate role in statutory construction. Given this lack of guidance, the Court's attempt to reaffirm its distrust of legislatures in *Osborne* may have resulted in an undesirable nod toward state judicial activism. *Osborne* seems to signal a greater tolerance of state-court interference in the legislative function, in effect allowing state judges to rewrite their legislatures' statutes.[36] Because the Supreme Court "invariably accepts the gloss the highest state court has placed on a state statute,"[37] a state court revision of a statute likely will remain uncorrected.

The reasoning of *Osborne* seems strangely at odds with Justice Scalia's opinion in *Oakes*, a case decided less than one year earlier. *Oakes* stands for the proposition that legislators should not be absolved from sloppy drafting by having a chance to amend a statute after it is challenged in court.[38] The *Osborne* majority states, without citing any authority, that legislators will be deterred from such careless drafting by the fear that courts will invalidate, rather than narrow, an overbroad statute.[39]

---

[31] L. TRIBE, *supra* note 10, at 1032.

[32] 92 U.S. 214 (1875).

[33] *Id.* at 221.

[34] 367 U.S. 203 (1961).

[35] *Id.* at 211.

[36] "The demarcation between 'statutory interpretation' or 'constitutional interpretation,' on the one hand, and judge-made law on the other is not a sharp line. Statutory interpretation shades into judicial lawmaking on a spectrum, as specific evidence of legislative purpose with respect to the issue at hand attenuates." P. BATOR, D. MELTZER, P. MISHKIN & D. SHAPIRO, HART AND WECHSLER'S THE FEDERAL COURTS AND THE FEDERAL SYSTEM 863 (3d ed. 1988).

[37] Monaghan, *supra* note 9, at 21–22. *See* L. TRIBE, *supra* note 10, at 1030 n.6, 1032 n.3.

[38] *Oakes*, 109 S. Ct. at 2639–40 (Scalia, J., concurring and dissenting). In Winters v. New York, 333 U.S. 507, 514 (1948), the Court held that a state court construction "puts these words in the statute as definitively as if it had been so amended by the legislature." Therefore, if *Oakes* stands for the proposition that *legislatures* should not be able to "put these words in the statute" after it is challenged, then it seems that *courts* should be prevented from doing the same thing.

[39] *Osborne*, 110 S. Ct. at 1702.

This reasoning is unrealistic. Given the political nature of judges in most states,[40] there is always the possibility of collusion between judges and legislators. Absent collusion, judges faced with an overbroad law may reasonably conclude that minor problems with a statute do not justify sending it back to the legislature for another round of drafting, hearings, and debate. Once state legislators are aware of this judicial propensity to narrow, rather than invalidate, laws, their incentives for careful drafting may begin to evaporate. Therefore, if the Court was as serious about the need for careful drafting as *Oakes* suggests, it should have required state courts to invalidate overbroad laws instead of adopting a saving construction, the only exception being when a clear legislative intent in accord with the saving construction can be discerned.[41]

It is ironic that a Supreme Court majority noted for judicial restraint would affirm the use of such broad power by state courts. The attempts to reconcile this contradiction are not convincing. In her plurality opinion in *Oakes*, Justice O'Connor noted that "the [legislative] amendment of a statute pending appeal to eliminate overbreadth is not different, in terms of applying the new law to past conduct, from a state appellate court adopting a limiting construction of a statute to cure overbreadth."[42] In his *Osborne* opinion, Justice White stated that the Court "has long respected the state Supreme Courts' ability to narrow state statutes so as to limit the statute's scope to unprotected conduct."[43] Justice O'Connor cited no support for her

---

[40] "The active participation of state judges in the policy process is much more taken for granted and much less controversial than the involvement of federal judges in the national government." Linde, *Observations of a State Court Judge*, in JUDGES AND LEGISLATORS TOWARD INSTITUTIONAL COMITY 117 (R. Katzmann ed. 1988). *See Built-In Lobby: Court Has Friends on the Hill*, Boston Globe, Sept. 26, 1990, at 1, col. 1 (describes close relationship between judges and legislators in Massachusetts).

[41] Justice Scalia has been a staunch opponent of judicial policymaking. In INS v. Cardoza-Fonseca, 480 U.S. 421, 452–53 (1987) (Scalia, J., concurring), he noted, "[j]udges interpret laws rather than reconstruct legislators' intentions. Where the language of those laws is clear, we are not free to replace it with unenacted legislative intent." *See* Johnson v. Transportation Agency, 480 U.S. 616, 670 (1987) (Scalia, J., dissenting) (cautioning the Court against "rewr[iting] the statute it purported to construe"). For other discussions of the use of legislative intent in statutory interpretation, see Zeppos, *Legislative History and the Interpretation of Statutes: Toward a Fact-Finding Model of Statutory Interpretation*, 76 VA. L. REV. 1295 (1990); Easterbrook, *The Role of Original Intent in Statutory Construction*, 11 HARV. J.L. & PUB. POL'Y 59 (1988); Aleinikoff, *Updating Statutory Interpretation*, 87 MICH. L. REV. 20 (1988); Starr, *Observations About the Use of Legislative History*, 1987 DUKE L.J. 371.

[42] *Oakes*, 109 S. Ct. at 2638.

[43] *Osborne*, 110 S. Ct. at 1702.

proposition, and the lone case cited by Justice White was *Ginsberg v. New York.*[44]

In *Ginsberg*, the Court upheld the conviction of a lunch counter owner for selling a "girlie" magazine to a minor in violation of New York law.[45] The defendant challenged the term "harmful to minors" in the statute as impermissibly vague.[46] *Ginsberg*, however, hardly supports either Justice White's or Justice O'Connor's propositions, since the New York Court of Appeals never really asserted its power to narrow an overbroad law. In fact, the defendant was denied leave to appeal to the New York Court of Appeals.[47] *Ginsberg* simply affirmed the New York court's finding in a similar case two years earlier[48] that the legislature's definition of "harmful to minors"[49] was sufficiently specific. Moreover, this definition was "virtually identical"[50] to the Supreme Court's definition of obscenity at that time.[51] In short, no narrowing construction was undertaken by the New York Court of Appeals in *Ginsberg*.

The lack of authority to support state court narrowing constructions might be excusable were it not for the Ohio Supreme Court's rather troubling actions in *Osborne*. Had Justice White and the rest of the *Osborne* majority done even a cursory examination of the Ohio child pornography statutes, they would have realized that the Ohio Supreme Court went well beyond ordinary statutory construction. In narrowing the law, the Ohio court directly contradicted the legislature's unmistakable intent that its laws have a broad application. The Ohio court's finding that the statute only applied to "a lewd exhibition [or] a graphic focus on the genitals"[52] was at odds with the statute's expansive definition of nudity:

---

44 390 U.S. 629 (1968).

45 The New York law made it unlawful "knowingly to sell . . . to a minor" a magazine "which depicts nudity . . . and which is harmful to minors." *Ginsberg*, 390 U.S. at 647.

46 *Id.* at 643.

47 *Id.* at 633.

48 Bookcase, Inc. v. Broderick, 18 N.Y.2d 71, 271 N.Y.S.2d 947, 218 N.E.2d 668 (1966).

49 The New York statute defined "harmful to minors" as representation of nudity which "(i) predominantly appeals to the prurient, shameful or morbid interest of minors, and (ii) is patently offensive to prevailing standards in the adult community as a whole with respect to what is suitable material for minors, and (iii) is utterly without redeeming social importance for minors." *Ginsberg*, 390 U.S. at 646.

50 *Id.* at 643 (quoting *Bookcase*, 18 N.Y.2d at 76, 271 N.Y.S.2d at 953, 218 N.E.2d at 672).

51 A Book Named "John Cleland's Memoirs of a Woman of Pleasure" v. Attorney General of Mass., 383 U.S. 413 (1966); Roth v. United States, 354 U.S. 476 (1957).

52 State v. Young, 37 Ohio St. 3d 249, 252, 525 N.E.2d, 1363, 1368 (1988). In his

> "Nudity" means the showing, representation, or depiction of human male or female genitals, pubic area, or buttocks with less than a full, opaque covering, or of a female breast with less than a full, opaque covering of any portion thereof below the top of the nipple, or of covered male genitals in a discernibly turgid state.[53]

The Ohio legislature's use of such a broad definition of "nudity" may have been motivated by a genuine concern about the dangers of child pornography. The legislature certainly could have drafted a narrower statute; it chose not to. For example, other provisions of the Ohio statute prohibit the creation, sale, distribution, or possession of "obscenity involving a minor"[54] and materials depicting a minor engaged in "sexual activity, masturbation, or bestiality."[55] Both provisions go well beyond the nebulous term of "nudity."

On the other hand, the three provisions of section 2907.323(A) prohibit photography of a minor who is not one's child or ward in a "state of nudity," consenting to nude photography of one's minor, and possessing or viewing nude photos of a minor who is not one's child or ward.[56] In each of these provisions, the term "state of nudity" is used; thus, the Ohio legislature clearly intended to distinguish between the terms "state of nudity," "obscenity involving a minor," and "sexual activity, masturbation or bestiality." If not, then all three provisions would be redundant. It is difficult to imagine that the legislature meant "lewd exhibition [or] graphic focus on the genitals" when it used the term "nudity." In fact, the word "lewd" appears nowhere in the statutory definition of any sex offense in the Ohio statute.[57] "Thus, when the Ohio Supreme Court grafted the 'lewd exhibition' test onto the definition of nudity," Justice Brennan wrote, "it was venturing into uncharted territory."[58] Moreover, accord-

---

Osborne dissent, Justice Brennan challenged the Ohio court's construction on two other grounds: the narrowed law was still overbroad because it covered "lewd exhibitions of *nudity*" rather than "lewd exhibitions of *the genitals*," and the terms "lewd" and "graphic focus" were impermissibly vague. *Osborne*, 110 S. Ct. at 1707 (Brennan, J., dissenting) (emphasis in original).

[53] OHIO REV. CODE ANN. § 2907.01(H) (Supp. 1989).

[54] *Id.* § 2907.321.

[55] *Id.* § 2907.322.

[56] *Id.* § 2907.323(A). Sections 2907.323(A)(1) & (2) both contain exception clauses for "proper purposes" which are identical to section 2907.323(A)(3). *See supra* note 21.

[57] *Osborne*, 110 S. Ct. at 1709 n.8 (Brennan, J., dissenting).

[58] *Id.* at 1710.

ing to Justice Brennan, the "graphic focus" element created by the Ohio court was "a stranger to obscenity regulation."[59]

In formulating its narrowing construction, the Ohio court cited no legislative history but merely noted that section 2907.323(A)(3)(a) contains an exception to the general prohibition against nude photos of children in cases of "proper purposes." According to the Ohio court, the statute allows the possession or viewing of nude photos "where that conduct is morally innocent."[60]

> Thus, the only conduct prohibited by the statute is conduct which is *not* morally innocent, i.e., the possession or viewing of the described material for prurient purposes. So construed, the statute's proscription is not so broad as to outlaw all depictions of minors in a state of nudity, but rather *only those depictions which constitute child pornography.*[61]

The circularity of this argument is obvious: the only unprotected activity is child pornography, which by definition seems unprotected by the Ohio statute. If the Ohio court's actions were as broad as this Essay has suggested, one is left wondering why a conservative majority of the Supreme Court was willing to sanction such expansive judicial power. Perhaps the Court was genuinely concerned about the dangers of child pornography. On the other hand, it may have followed tradition and deferred to a state supreme court's interpretation of a state statute. Or it is even possible that the *Osborne* majority simply was not aware of what the Ohio court had done.

Whatever the Supreme Court's motivation, this affirmation, and perhaps expansion, of state court power is troublesome for several reasons. First, once narrowing construction crosses the boundary into legislating, as seemed to happen in *Osborne*, a court is making rather than interpreting the law.[62] Such quasi-legislative action seems to violate the explicit separation of

---

[59] *Id.* at 1711.

[60] State v. Young, 37 Ohio St. 3d 249, 252, 525 N.E.2d 1363, 1367 (1988).

[61] *Id.* at 1367–68 (emphasis added). This reliance on "proper purposes" raises vagueness problems. As Justice Brennan noted in his dissent, "[w]hat is a permissible 'other proper purpose'? What about photos taken for one purpose and recirculated for other, more prurient purposes? The 'proper purposes' standard appears to create problems analogous to those this Court has encountered in describing the 'redeeming social importance' of obscenity." *Osborne*, 110 S. Ct. at 1706 n.2 (Brennan, J., dissenting).

[62] To solve this separation of powers problem, Dimond and Jeffrey propose that state courts adopt a referee role in intergovernmental disputes, leaving ultimate power to the legislatures. Dimond & Jeffrey, *An Appropriate Role for State Courts in Intergovernmental Disputes: The Referee Model*, 32 WAYNE L. REV. 51 (1985).

powers provisions that most states have in their constitutions.[63] As Justice Kogan of the Florida Supreme Court noted in a case involving an overbreadth challenge to an obscenity statute:

> [C]ourts may not engage in the essentially legislative act of varying actual intent or reading new elements into a statute . . . which would violate the separation of powers doctrine . . . . [W]hen the subject statute in no way suggests a saving construction, we will not abandon judicial restraint and effectively rewrite the enactment. The Florida Constitution requires a certain precision defined by the legislature, not legislation articulated by the judiciary.[64]

The danger of a state supreme court essentially rewriting an overbroad law is magnified by the fact that separation of powers is traditionally policed by the courts.[65]

A second objection to broad judicial power is that state courts lack the institutional competence[66] to decide such issues. Legislatures have significantly more resources and expertise than the courts to investigate a problem and draft an appropriate solution.

A final objection is that state judges are less likely to reflect the will of the public, making any judicial lawmaking inherently

---

[63] For example, a provision in the Arizona constitution which is similar to provisions in many other state constitutions, states:

> The powers of the government of the State of Arizona shall be divided into three separate departments, the Legislative, the Executive and the Judicial; and, except as provided in this Constitution, such departments shall be separate and distinct, and *no one of such departments shall exercise the powers properly belonging to either of the others.*

ARIZ. CONST. art. III (emphasis added). Thirty-four states have explicit separation of powers provisions in their constitutions. Browde & Occhialino, *Separation of Powers and the Judicial Rule-Making Power in New Mexico: The Need for Prudential Constraints*, 15 N.M. L. REV. 474 app. A (1985). In the other states, as in the federal system, an implied separation of powers doctrine is derived from the creation of separate legislative, executive, and judicial branches. *Id.* at 408 n.3. Though the Massachusetts constitution has a separation of powers provision, MASS. CONST. art. 30, and the Ohio constitution does not, this does not appear to have influenced the Court's decisions in either *Oakes* or *Osborne*.

[64] Stall v. Florida, No. 74,020, 74,390 (Fla. Oct. 11, 1990) (LEXIS, States library, Fla. file) (Kogan, J., dissenting) (quoting Brown v. State, 358 So.2d 16, 20 (Fla. 1978)).

[65] *See, e.g.*, Mistretta v. United States, 488 U.S. 361 (1989).

[66] "As a popularly elected body, the [state] legislature is in a position to tap the thinking of its constituency and has the resources to secure data generally not available to the courts." *In re* Asbestos Litig., 829 F.2d 1233, 1240 (3d Cir. 1987). Judge Hans A. Linde of the Oregon Supreme Court argues, however, that state judges are just as capable of deciding policy issues as legislators since many judges have spent time in the state legislature. He notes that "the judiciary is perceived as being a more professional and permanent institution than state legislatures, which meet only intermittently and have relatively weak institutional structure . . . . [T]he smaller geographic distance of state judges from their state capitals makes it relatively easy for them to stay in touch with legislative activities." Linde, *supra* note 40, at 118.

undemocratic, or at least less democratic than statutes passed by a legislature. As Professor Tribe notes, "[T]he doctrines of overbreadth and vagueness capture the essence of a demand that, *in close cases, government must leave speech ample room to breathe*. How best to do that is properly left to the majoritarian branches . . . ."[67] The idea that state judges are more accountable to the public because they are popularly elected is becoming increasingly outdated, as more states move to merit selection of judges.[68] And in the thirty-nine states which do elect state supreme court judges,[69] little public attention is given to judicial races.[70] Professors Solimine and Walker note that "few, if any, state judges are defeated because of their decisions favoring federal rights."[71]

Even when state courts are exercising legitimate authority in narrowing overbroad statutes, such saving constructions must not violate due process. The Court has long noted that statutes may be applied to conduct by a narrowing construction "provided such application affords fair warning to the defendant[ ]."[72] Such constructions also cannot be "unexpected" or "unforeseeable."[73] In *Osborne*, the majority paid only lip service to Osborne's due process interests when it reversed his conviction. Justice White missed the point when he noted that the trial court should have instructed the jury that "nudity" meant "lewdness."[74] The real issue is not whether Douglas Osborne was convicted under a sufficiently narrow statute, but whether he had fair warning when he photographed his daughter that his actions were not constitutionally protected.[75] As one commen-

[67] L. TRIBE, *supra* note 10, at 1039 (emphasis in original).

[68] Solimine & Walker, *State Court Protection of Federal Constitutional Rights*, 12 HARV. J.L. & PUB. POL'Y 127, 136 n.4 (1989). *Contra* Dimond & Jeffrey, *supra* note 62, at 51 n.42 ("state courts are subject to many more democratic restraints in interpreting state constitutions [than the federal courts in interpreting the U.S. Constitution] because most state judges are elected and most state constitutions are more readily amended").

[69] *For Want of Recognition, Chief Justice is Ousted*, N.Y. Times, Sept. 28, 1990, at B16, col. 3.

[70] Solimine & Walker, *supra* note 68, at 136. According to an expert in judicial elections, "[w]hat nobody counted on was that most voters don't follow state supreme court elections very closely and when they go into the voting booth they often pick a name that sounds familiar, usually the more common name." N.Y. Times, *supra* note 69, at B16, col. 3.

[71] Solimine & Walker, *supra* note 68, at 136.

[72] Dombrowski v. Pfister, 380 U.S. 479, 491 n.7 (1965).

[73] Marks v. United States, 430 U.S. 188, 192 (1977).

[74] *Osborne*, 110 S. Ct. at 1703–05.

[75] The fact that both the Massachusetts and Ohio statutes went well beyond other child pornography statutes may strengthen this due process argument. *See* New York

tator noted about *Oakes*, "[such] a subsequent and unpredictable act of the legislature to alter the defendant's legal defense . . . is unduly reminiscent of an ex post facto law."[76] If the thesis of this Essay is correct, then a judicial narrowing of an overbroad law has similar ex post facto characteristics.

Just as the *Osborne* majority did not look at fair warning to Douglas Osborne, it did not consider the potential chilling effect of subsequently narrowed laws on the protected speech of others.[77] One who innocently photographs her "naked one-year-old running on a beach or romping in a wading pool,"[78] or visits a Robert Mapplethorpe exhibit that features photos of nude children[79] is no more on notice that her conduct is unprotected when courts amend the laws after the fact than when legislatures do so. Since much of the protected activities in *Oakes* and *Osborne*—for example, family photos and nude sunbathing—are "private and unpublicized,"[80] it is difficult to ascertain the precise chilling effect[81] on "the prudent, the cautious and the circumspect" members of society.[82] Though it may be a "Herculean task to draft a statute that . . . survive[s] scrutiny for vagueness and overbreadth,"[83] an individual without notice of

---

v. Ferber, 458 U.S. 747, 751 (1982) (New York law proscribed "promoting a sexual performance" by a child).

[76] Recent Development, *First Amendment Overbreadth Doctrine—Massachusetts v. Oakes*, 25 HARV. C.R.-C.L. L. REV. 221, 235–36 (1990).

[77] Professor Tribe has noted that in other first amendment cases Justice White has expressed "skepticism . . . toward the reality and significance of the deterrence caused by an overbroad law." L. TRIBE, *supra* note 10, at 1026.

[78] Commonwealth v. Oakes, 401 Mass. 602, 605, 518 N.E. 2d 836, 838 (1988).

[79] An amicus brief filed by the Law and Humanities Institute in *Oakes* noted that a large portion of contemporary art featuring non-pornographic nude photography, films, and paintings could be prosecuted under the unamended Massachusetts statute. *Oakes*, 109 S. Ct. at 2643 n.3. (Brennan, J., dissenting). In *Osborne*, Justice Brennan noted that "visitors to an art gallery might find themselves in violation of the [Ohio] law," which prohibited mere viewing of nude photographs of children. *Osborne*, 110 S. Ct. at 1712 n.13 (Brennan, J., dissenting).

[80] Recent Development, *supra* note 76, at 238.

[81] But in Board of Airport Comm'rs of Los Angeles v. Jews for Jesus, Inc., 482 U.S. 569, 575–76 (1987), the Supreme Court struck down, rather than narrowed, a rule against first amendment activities at Los Angeles International Airport, noting, "[I]t is difficult to imagine that the resolution could be limited by anything less than a series of adjudications, and the chilling effect of the resolution on protected speech in the meantime would make such a case-by-case adjudication intolerable." *See* Arnett v. Kennedy, 416 U.S. 134, 231 (1974) (Marshall, J., dissenting) ("For every employee who risks his job by testing the limits of the statute, many more will choose the cautious path and not speak at all"); Keyishian v. Board of Regents, 385 U.S. 589, 604 (1967) ("When one must guess what conduct or utterance may lose him his position, one necessarily will 'steer far wider of the unlawful zone.'" (quoting Speiser v. Randall, 357 U.S. 513, 526 (1958))).

[82] Spears v. State, 337 So.2d 977, 980 (Fla. 1976).

[83] Comment, *supra* note 3, at 605.

the law's scope should not be penalized because of statutory imprecision. As the Court noted over a century ago, "[i]t would certainly be dangerous if the legislature could set a net large enough to catch all possible offenders, and leave it to the courts to step inside and say who could be rightfully detained, and who should be set at large."[84] It would be strange constitutional law indeed if Justice Scalia's *Oakes* reasoning was reduced to an artificial distinction between legislatures and courts.

The solution, then, is for state courts to invalidate overbroad laws when they are "rotten at [their] very root,"[85] rather than "venturing into uncharted territory"[86] by rewriting a law.[87] In *Osborne*, voiding the child pornography law was the only legitimate option, given the Ohio legislature's clear intention to draft a broad statute. The Court's current preference, however, is for reconstructive surgery on a statute rather than simple invalidation,[88] usually by severing the overbroad provisions from the statute.[89] This second option, however, is impractical when the offending provision is closely interwoven with the rest of the statute.[90] This was the case in *Osborne*. For a court to remove the term "nudity" from section 2907.323(A)(3) effectively would have rendered the rest of the statute meaningless.[91] When severance is possible, however, it is likely to contradict the legislature's purpose.[92] Since legislation is often the result of subtle compromises, one part cannot be severed without affecting the integrity of the entire law. Judicial invalidation, on the other

---

[84] United States v. Reese, 92 U.S. 214, 221 (1876).

[85] L. TRIBE, *supra* note 10, at 1029.

[86] *Osborne*, 110 S. Ct. at 1710 (Brennan, J., dissenting).

[87] In a recent Florida case, an intermediate court found the state's law proscribing "sexual conduct" with a minor overbroad but construed the statute to apply only to "lewd or lascivious conduct." Schmitt v. State, 563 So.2d 1095, 1099–1100 (Fla. Dist. Ct. App. 1990). There was no evidence that the legislature intended such a construction.

[88] L. TRIBE, *supra* note 10, at 1027.

[89] *See* Brockett v. Spokane Arcades, Inc., 472 U.S. 491 (1985) (reversing a decision to invalidate rather than sever a Washington obscenity statute).

[90] *See* Village of Schaumburg v. Citizens for a Better Environment, 444 U.S. 620 (1980) (striking down solicitation ordinance since severability was impossible); American Booksellers Ass'n v. Hudnut, 771 F.2d 323, 332 (7th Cir. 1985), *aff'd mem.*, 475 U.S. 1001 (1986) (striking down pornography ordinance despite "strong severability clause," since no "excision of particular terms could save it").

[91] In *Hill*, the Court opted against severing unconstitutional provisions, noting, "[I]t is doubtful that even a remarkable job of plastic surgery upon the face of the ordinance could save it." Houston v. Hill, 482 U.S. 451, 469 (1987) (quoting Shuttlesworth v. Birmingham, 394 U.S. 147, 153 (1969))

[92] Judicial severance is also less likely to deter legislatures from careless drafting. *See* text accompanying note 40–41.

hand, allows the legislature to reexamine the entire problem and redraft a comprehensive statute.

III. CONCLUSION

Greater authority to narrow overbroad laws can be a dangerous weapon in the hands of even the most well-intentioned state courts. At best, state judges will substitute their own views for those of the legislators, thus crossing the line into lawmaking. At worst, judicially narrowed speech laws will violate due process and chill protected expression.

While the significance of *Osborne* may be subtle, the decision has important implications in the area of free speech—an area in which even minor decisions can have sweeping impacts. Whether state courts embark on substantial rewriting of overbroad laws in the wake of *Osborne* remains to be seen. In the first amendment context, however, the danger is that legitimate, protected speech will be chilled by a Sword of Damocles, whose danger "is that it hangs—not that it drops."[93] Such a danger can only be averted by having both state courts and legislatures tread lightly in this sensitive area.

—Christopher P. Lu

---

[93] Arnett v. Kennedy, 416 U.S. 134, 231 (1974) (Marshall, J., dissenting).

RECENT PUBLICATIONS

ON THE LAW OF NATIONS. By *Daniel Patrick Moynihan*. Cambridge, Mass.: Harvard University Press, 1990. Pp. 177, notes, index. $22.50 cloth.

In his latest book, Senator Daniel Patrick Moynihan (D-N.Y.) has set out to accomplish what at times seems to be a quixotic task. Seeing in the recent foreign policy of the United States an ominous trend toward disregarding the norms of international law, he attempts to rekindle American enthusiasm for the "Law of Nations." Moynihan's analysis proceeds in two parts. First, he traces the historical roots of the United States's traditional faith in international law. He then turns to the present state of international relations in order to examine the challenges that nation-states will face in the twenty-first century and the rules of international conduct with which they might meet these challenges.

This is not an abstruse international legal manual meant to be laboriously parsed to extract the principles of international law. As Moynihan admits, "I . . . have little to say on the subject of international law itself save that it exists" (p. 13). Nonetheless, to the "realist" school that tends to view international law as nothing more than a doctrinal apology for an anarchic world system in which what is called international law is in fact the law of the victor, Moynihan has an extended answer. He writes that "the long twilight struggle [of the believers in international law] is ending; we appear to have prevailed" (pp. 13–14). For the rest of the book he treats those skeptics to a sometimes rambling, but unfailingly interesting, discourse on the origins and development of the traditional American enthusiasm for and support of the law of nations, and ends with a warning about the consequences of present American skepticism regarding international law (p. 177).

He starts with the Framers of the Constitution, who, he notes rather wistfully, "were close to ancient things" (p. 16). Their belief in a law of nations derived from the classical Greek and Roman concept (later amplified in medieval times by Grotius) of a law of nature (p. 16). This concept had evolved in Europe in the context of the relationships of nations to include a "code of public instruction, which defines the rights and prescribes the duties of nations, in their intercourse with each other. The faithful observance of this law is essential to national character, and to the happiness of mankind" (p. 15).[1] Thus, during the American Revolution, the Continental Congress "professed obedience" to the law of nations (p. 15) and eleven years later made the law of nations an exclusively federal concern by granting original district court jurisdiction over "all causes where an alien sues for a tort only [committed] in violation of the law of nations . . . ."[2] As Moynihan makes clear, American scholars and jurists early on mingled the natural law concepts underlying both democracy and international law (p. 19). The liberal values that underlay the Bill of Rights also informed the American conception of international order. This conception of international order was most forcefully articulated by Woodrow Wilson in 1919. Wilson expressed a vision of transcendant values of universal liberty and self-determination in speeches he made across the country as he tried to muster support for the Treaty of Versailles that was to form the basis of the Covenant of the League of Nations (pp. 52–53).

Wilson's historic struggle to gain the authority with which to implement his global vision introduces a major theme running throughout the book—the importance of the relationship between Congress and the President in the conduct of foreign policy. President Wilson had forgotten—or had wilfully chosen to ignore—what Professor Wilson had recognized in *Congressional Government*: the supremacy of the Congress (and especially the power of the Senate with regard to treaties) in the foreign policy-making process (pp. 44–45).[3] His visionary plan for international government—the League of Nations—needed the support of powerful Senators such as Henry Cabot Lodge of Massachusetts, who was both Republican floor leader and chairman of the Committee on Foreign Relations (p. 48). Instead of trying to assuage their worries about the Treaty's potential for dragging the United States into war by guaranteeing the territorial integrity of every League member, Wilson engaged in petty name-calling, branding the Senators opposing him, particularly Lodge, "contemptible . . . narrow . . . selfish . . . poor little minds that never get anywhere but run round in a circle

---

[1] J. KENT, 1 COMMENTARIES ON AMERICAN LAW 1 (1826).

[2] Judiciary Act of 1789, 1 Stat. 73, 77 (codified as amended at 28 U.S.C. 1350 (1988)). For a modern application of this rarely used cause of action, see Filartiga v. Pena-Irala, 630 F.2d 876 (2d Cir. 1980).

[3] W. WILSON, CONGRESSIONAL GOVERNMENT (1885).

and think they are going somewhere" (pp. 50–51).[4] As Moynihan notes, "Wilson proved there is such a thing as a man too proud to conciliate. . . . The Treaty of Versailles was defeated before it was negotiated" (p. 46).

This lesson was well-learned by Franklin D. Roosevelt, who, with the help of Congress, accomplished what Wilson had set out to do: create an international organization—the United Nations—with a viable decisionmaking and enforcement mechanism in the form of the Security Council (p. 78). "Roosevelt took Congress every step of the way. The Congress was no more 'enlightened' in 1944 than it had been in 1919. If anything, less so" (p. 71). Yet Roosevelt was adept at manipulating Congress, as when he avoided congressional opposition to the "destroyer for bases" deal with Great Britain through a misleading legal opinion by Attorney General Robert H. Jackson regarding the Neutrality Act of 1917 (p. 71).[5]

Yet Moynihan acknowledges that getting the Executive and Congress to work together is only the first step toward the universal acceptance of international law. As Moynihan notes, too often "[l]aw is confused with force" (p. 132). The concept of law as nothing more than power was formulated best by Thucydides, who quoted the victorious Athenians telling the Melians, "[t]he strong do what they can and the weak suffer what they must" (p. 102).[6] This view became widespread during the Cold War era. Initially developed in reaction to the utopianism of the inter-war years, and spurred on by the memory of Nazi Germany and the disillusionment of the Cold War, the "realist" school of international relations became ascendant.[7] These realists "looked upon international law as the delusion of the well-intentioned but inexperienced" (p. 131). Moynihan writes in his inimitable style that "in the annals of forgetfulness there is nothing quite to compare with the fading from the American mind of the idea of the law of nations" (p. 99). That is, with the advent of the Cold War and the onset of superpower

---

[4] R. BYRD, *The Senate: 1789–1989*, in 1 ADDRESSES ON THE HISTORY OF THE UNITED STATES SENATE 423 (M. Hall ed. 1988).

[5] Attorney General Jackson inserted a comma into the 1917 Neutrality Act so that it would be read only as forbidding the United States from providing its allies with vessels built with the specific intent that such vessels would be delivered to a belligerent nation. Without the comma, the clause would have been read as forbidding the provision of any vessel.

[6] THUCYDIDES, THE PELOPONNESIAN WARS 351–53 (R. Crawley trans. 1982).

[7] *See* E. CARR, THE TWENTY YEARS CRISIS (1981); H. MORGENTHAU, POLITICS AMONG NATIONS (1985).

hostility, the realists's conception of international law became the dominant influence in the United States's foreign policy.[8]

Then came Mikhail Gorbachev. Moynihan's message of hope regarding the future of international law paradoxically draws its most significant support in the late twentieth century from the President of the Soviet Union. Mikhail Gorbachev's speech to the United Nations on December 7, 1988, proposed significant arms reduction and an end to the Cold War; "the world," prophesies Moynihan, "had entered a new stage" (p. 81). Essential to Moynihan's thesis that the future for international law is bright seems to be the assumption that a respect for international law thrives when all polities converge toward the ideal of political liberalism. Though he never explicitly defines the term, Moynihan's concept of political liberalism in its international context includes a healthy respect for international law. Thus perestroika and the dismantling of the authoritarian regimes of Eastern Europe are essential to this book's message of hope. As he puts it, "[p]eople change their minds" (p. 82). This seemingly simplistic notion forms the basis of Moynihan's serious hypothesis that in the last decade of the twentieth century we are witnessing the irreversible decay of ideological differences, and are shifting toward a "functional rationality" that bodes well for the future stability of international law (p. 83).

Moynihan suggests that re-achieving international respect for international law means that all nations must adhere to such respect consistently. The United States, in 1985, announced that it would not be subject to the jurisdiction of the International Court of Justice ("ICJ") (p. 146). In 1986, the ICJ found that various activities the United States had undertaken against the government of Nicaragua constituted a violation of international law (pp. 146–47). "Pacta sunt servanda!" (agreements must be honored) thunders Senator Moynihan, echoing what President Gorbachev had said in another context (p. 98). "A political culture from which the idea of international law has largely disap-

---

[8] The author at p. 133 quotes Dr. Jeane Kirkpatrick, President Ronald Reagan's ambassador to the United Nations, stating that "[w]e cannot permit . . . ourselves to feel bound to unilateral compliance with obligations which do in fact exist under the Charter, but are renounced by others . . . ." Kirkpatrick, *Law and Reciprocity*, 78 PROCEEDINGS OF THE ANNUAL MEETING, AMERICAN SOCIETY OF INTERNATIONAL LAW 59 (1984). Moynihan also notes that CIA Director William Casey came to Moynihan's office and acknowledged that he had broken the law by deceiving the Senate about the Iran-Contra affair, but stressed that the law to which he was referring was the Intelligence Oversight Act. Casey told Moynihan that "he had no interest in treaty law" (p. 197 n.38).

peared places its initiatives in jeopardy" (p. 148)—a brave statement that gains its power from the present international context. Moynihan argues that one of the fundamental causes for the failure of President Reagan's policy toward the Nicaraguan Contra rebels was that "men and women responsible for national security affairs were either ignorant or contemptuous of international law" (p. 121).

What he does not make as clear, however, is the extent of the controversy concerning the power of the President to make treaties and executive agreements in areas in which both Congress and the Executive have authority.[9] That is, where neither the Constitution nor explicit congressional delegation of authority allows the President to make international obligations which are binding upon the United States, there exists a zone of concurrent authority. This was recognized as far back as the debates between Alexander Hamilton (writing as Pacificus) and James Madison (writing as Helvidus). Hamilton argued that by granting the President the executive power in Article II, the Constitution granted the President the power over foreign affairs that was implied in the idea of executive power, "except as expressly modified in the Constitution."[10] Madison believed that the President only had the power to make treaties and appointments that was given to him expressly by the Constitution and those powers implied in the President's role as Commander in Chief, with the latter powers "not to be extravagantly construed."[11]

While Moynihan obviously tends toward the Madisonian view of executive power (pp. 173–74), the Hamiltonian view certainly has its adherents. The Supreme Court, for example, has often upheld the President's authority in foreign affairs, most famously in *United States v. Curtiss-Wright Export Corp.*,[12] where it stated that "[t]he President is the sole organ of the nation in its external relations, and its sole representative with foreign nations."[13] However, as one scholar has noted, if the President

---

[9] *See* Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579 (1952). Justice Jackson's famous concurrence outlined this "zone of twilight in which [the President] and Congress may have concurrent authority, or in which its distribution is uncertain. . . . In this area, any actual test of power is likely to depend on the imperatives of events . . . rather than on abstract theories of law." *Id.* at 637 (Jackson, J., concurring).

[10] L. HENKIN, CONSTITUTIONALISM, DEMOCRACY, AND FOREIGN AFFAIRS 22 (1990).

[11] *Id.*

[12] 299 U.S. 304, 319 (1936).

[13] *Id.* at 319 (quoting 6 ANNALS OF CONGRESS 613 (1800) (Speech of Rep. John Marshall in the House of Representatives on March 7, 1800)). *See also* United States v. Pink, 315 U.S. 203 (1942); Chicago & Southern Air Lines v. Waterman S.S. Corp.,

insists on his authority, "Congress can usually prevail, in constitutional principle and in governmental practice, if only because it holds the purse strings."[14] This putative separation of powers thus results in much wrangling between these two government branches.

While the internal institutional problems facing the United States in formulating its foreign policy necessarily increase the uncertainty of its position regarding international law, there are numerous other barriers to the development of universal recognition of international law. Moynihan stresses the difficulty of the enforcement of international legal norms in an anarchic world system rife with divisions based on racial, religious, and ethnic differences (pp. 158–59). Another problem, which Moynihan does not address directly, is that even as the stern ideological barriers between the superpowers seem to fade away, there remains an asymmetrical distribution of power among nations that produces different national dispositions toward international law. As the International Court of Justice recognized some time ago,

> [t]here is prevalent in the world today a widespread questioning of . . . contemporary international law. This feeling is based on the view that . . . the greater part of international law is the product of European imperialism and colonialism and does not take sufficient account of the completely changed pattern of international relations which now exists.[15]

Moynihan argues that "a state that finds itself tempted by self-interest to erode traditional norms may in time regret its conduct" (p. 149). This reliance on an enlightened self-interest to foster respect for international law is problematic. The Cold War is probably over, at least for the time being. The dismantling of authoritarian regimes in most parts of the world will probably result in the easing of military tensions. However, we have far to go before reaching the safe haven of shared international norms and a stable international regime that abides by those

333 U.S. 103 (1948). More recently, the Court upheld an expansive exercise of presidential foreign affairs power in Dames & Moore v. Regan, 453 U.S. 654 (1981), which dealt with President Carter's agreement ending the Iran Hostage crisis and establishing the Iran-United States Claims Tribunal to arbitrate all unsettled claims between the nationals of both countries.

[14] L. HENKIN, *supra* note 10, at 31.

[15] Barcelona Traction, Light and Power Co. (Belgium v. Spain), 1970 I.C.J. 3, 248–49 (Feb. 5, 1970) (quoting S. ROSENNE, 1 THE LAW AND PRACTICE OF THE INTERNATIONAL COURT 17–18 (1965)).

norms. While Moynihan has faith in the value of rational self-interest on the part of states to solve the problem of the lack of an enforcement mechanism in international law, his own description of American domestic problems in formulating its international legal positions highlights the ubiquitous difficulties sovereign states face in achieving coherent approaches to foreign policy.

Moynihan's argument that the realist conception of international law should yield to one that encompasses a set of normative principles which condition and constrain state behavior ultimately requires a leap of faith, notwithstanding intimations of rational self-interest. Though this may detract from the practical implications of his thesis, his book remains valuable for its wealth of political and historical anecdotes, as well as for its redeeming vision of the future of the international state system.

—M. Sharmini Mahendran

AGENCY UNDER STRESS: THE SOCIAL SECURITY ADMINISTRATION IN AMERICAN GOVERNMENT. By *Martha Derthick.* Washington, D.C.: The Brookings Institution, 1990. Pp. xii, 226, index. $32.95 cloth; $12.95 paper.

The Social Security Administration ("SSA") was established in 1935 to run the new system of old age and survivors' insurance introduced by President Franklin D. Roosevelt. It gradually developed into one of the federal government's largest and most respected bureaucracies: "the brightest ornament in the ever-expanding federal establishment," "the best, the elite, a model of what government could do for the citizens it serves."[1] During the 1970's and 1980's, however, the SSA became involved in two debacles of monumental proportions that sharply lowered the agency's prestige and raised serious questions about the capacity of a centralized federal bureaucracy to handle complex tasks efficiently and fairly. Martha Derthick's *Agency Under Stress* seeks to analyze these failures and to draw from them larger lessons about the importance of anticipating at the policymaking stage the kinds of administrative problems that later may arise in implementing new programs.

---

[1] Johnson, *Days of Endless Struggle, Drowning in a Sea of Paper*, Washington Post, Mar. 27, 1977, at A15, col. 2.

The SSA's troubles, Derthick explains, began in 1974 when the agency took over responsibility from the states for financing and administering need-based aid to low-income disabled, blind, and aged persons, under a new program known as Supplemental Security Income ("SSI") (p. 5). In terms of policy, the creation of SSI represented a significant advance insofar as it recognized the needs of groups that far too often had been ignored by the federal government. In terms of administration, however, the implementation of SSI was a disaster. Much of the data transferred from the states to the SSA was incomplete or inaccurate, and the computer systems developed by the SSA to aid in processing applications and payments proved to be inadequate. Many who should have received benefit checks never got them, while others were mailed payments for which they were ineligible. Frequently the benefit amounts were either too low or too high. Attempts to correct these problems were often unsuccessful because the computer system became overloaded with inquiries and crashed repeatedly. The SSA field offices were mobbed with complaints, backlogs mushroomed, field representatives put in close to two million hours of overtime, and agency morale suffered (pp. 26–31).

The SSA was only beginning to recover from this crisis when a new problem began to loom on the horizon. Since 1956, disabled workers had been eligible for the benefits they normally would have received at age sixty-five (p. 33).[2] In the mid-1970's, the number of disability insurance beneficiaries began to skyrocket, in part due to statutory changes in the eligibility requirements (pp. 33–34). Alarmed by such unprecedented growth in the program and concerned about Social Security's long-term financial viability, in 1980 Congress required the Secretary of Health and Human Services to institute a periodic review of the eligibility of all beneficiaries of disability insurance who had not been judged to be permanently disabled.[3] The force of this directive was given additional impetus by the arrival of the administration of President Ronald Reagan, which was dedicated to eliminating "waste and fraud" in government, and by

---

[2] The actual amendment in 1956 limited disability insurance eligibility to workers over age 50; a further amendment in 1960 abolished the age requirement. *See* M. DERTHICK, POLICYMAKING FOR SOCIAL SECURITY 431 (1979).

[3] Social Security Disability Amendments of 1980, Pub. L. No. 96-265, § 311, 94 Stat. 460.

the issuing of internal reports suggesting that as many as one in every five beneficiaries might be ineligible (pp. 34–37).

By April 1982, the SSA had reviewed 405,000 cases and had terminated benefits in about forty-seven percent of them (p. 42). While these results appeared to confirm Reagan administration suspicions, questions about the fairness of the termination decisions plagued the review process. Stories of obviously disabled persons who had lost their benefits appeared continually in the media (p. 169). Terminations were frequently reversed by administrative law judges ("ALJs"),[4] and the SSA's criteria for termination were challenged in the federal courts. In the early yet important case of *Patti v. Schweiker* (p. 138),[5] the Ninth Circuit held that the SSA had to show that a beneficiary's medical condition had changed in order to claim that he or she was no longer disabled.[6] However, the SSA disregarded this and other rulings, choosing to pursue a policy of "nonacquiescence," under which it continued to deny benefits to other similarly situated recipients and to litigate suits over termination criteria, not only in circuits that had not yet ruled on the proper standard, but even in circuits that had already decided previous cases against the SSA (pp. 139–42).

Increasingly, the SSA found itself at odds not only with the federal courts, but also with other members of the executive branch. The Solicitor General had refused to appeal *Patti v. Schweiker* to the Supreme Court (p. 145), and U.S. Attorneys such as Rudolph Giuliani in New York declined to defend the government's position in subsequent cases.[7] The Secretary of Health and Human Services even became the target of a lawsuit by the Association of Administrative Law Judges, charging that judges who did not deny appeals at a certain rate had been singled out for review, thereby compromising their independence.[8]

---

[4] ALJs reversed the decision to terminate benefits in 92,000 of 152,000 appeals heard during the period from February 1982 through September 1983. M. Derthick, *The Plight of the Social Security Administration*, in SOCIAL SECURITY AFTER FIFTY 101, 111 (E. Berkowitz ed. 1987).

[5] 669 F.2d 582 (9th Cir. 1982).

[6] *Patti v. Schweiker* actually involved a termination of SSI benefits that took place prior to the formal disability review ordered by Congress in 1980, but its holding was applicable to subsequent terminations. *See, e.g.*, Lopez v. Heckler, 713 F.2d 1432 (9th Cir. 1983) (p. 140).

[7] *Disability Reviews Spur Legal "Crisis"*, N.Y. Times, Sept. 9, 1984, at 38, col. 1.

[8] Association of Administrative Law Judges v. Heckler, 594 F. Supp. 1132 (D.D.C. 1984).

Gradually, the SSA began to retreat from its policy of non-acquiescence, agreeing in 1985 to instruct ALJs to follow the criteria of the relevant circuit, although the agency continued to follow its own stricter standards for review (p. 148). Finally, in 1988, the SSA published new regulations under which it agreed to observe judicially decreed standards of review, unless an actual appeal was planned (p. 149). In the meantime, however, thousands of recipients had endured the hardship of losing their benefits and having to go to court to regain them, and the SSA saw its one-time prestige as a professional, non-political agency dissipated in endless court battles (pp. 150–51). In *Agency Under Stress*, Derthick argues that these two episodes should not be written off as mere bureaucratic bungles resulting from simple computer glitches or the over-zealousness of administrators. Rather, for Derthick the problems involved in the implementation of both SSI and disability review are symptomatic of much more profound administrative issues that are too often ignored in the policymaking process.

In the case of SSI, for instance, Derthick suggests that few in the executive branch, or even in the SSA itself, adequately estimated the magnitude of the project the agency was undertaking. No one seems to have considered sufficiently the fundamental differences between administering an entitlement program such as social security, from which judgment calls were largely absent except in the case of disability determinations, and a means-tested program such as SSI, in which every case required a judgment about eligibility (pp. 48, 186–87). Too many people assumed that the SSA, with its record of professionalism and its mastery of contemporary computer technology, had the personnel and the expertise to tackle virtually any task (pp. 190–91).

The SSA's mission in creating SSI was also complicated by the way Congress designed the authorizing legislation. Originally, the establishment of SSI was based on the belief that the federal government could distribute benefits more rationally, equitably, and efficiently than could the states (pp. 68–70). But Congress required the states to maintain the benefit levels of all current recipients of state aid by threatening to withhold Medicaid grants from non-complying states. At the same time, however, Congress encouraged the states to transfer administration of these mandatory supplements to the SSA to maintain the agency's overall control (pp. 73–74). These provisions in the law thus had the paradoxical effect of building into the system the very inconsistencies and complications that federal administration was supposed to eliminate.

While the problems of SSI were due in large part to a failure to assess accurately the complexity of the project and the capacities of the SSA, the disasters of disability review resulted from a failure to listen. The SSA itself acknowledged the need for a more vigorous review program, but it pleaded with Congress to leave the agency "the flexibility to develop the most cost-effective program for re-examining cases where improvement seems likely" (p. 83).[9] Nevertheless, Congress set as a requirement what the SSA had announced only as a goal—the review of all non-permanent disability cases at least every three years. Similarly, Health and Human Services officials voiced their doubts concerning the cost savings projected by David Stockman's Office of Management and Budget ("OMB"), and suggested that ALJs and the courts would react negatively to "what they perceive to be a tightening of the program beyond the requirements of the law" (pp. 61–62).[10] Intent on finding budget savings, the Reagan administration ignored these concerns and forged ahead with the review.

These brief summaries can offer only a glimpse of Derthick's subtle, wide-ranging, and intensively researched analysis of why SSI and disability review turned into disasters for the SSA. Derthick systematically examines not only the causes of the SSA's own miscalculations, but the impact of the executive, legislative, and judicial branches on the SSA's administration of SSI and disability review. She does so in an effort to illuminate how the habits and interests of each of these bodies can promote administrative gridlock. In her chapters on the Presidency, for instance, Derthick notes how a President's typical policymaking impulses to institute rapid and sweeping reforms and to rationalize existing programs are often at odds with the administrative goals of minimizing budget expenditures and reducing personnel. She also points out how the distance between the Presidency and the field level of administration can lead to fatal errors in judgment about the capabilities of agency personnel, and how the practice of filling agency positions based on

---

[9] *Disability Insurance Legislation: Hearings Before the Subcomm. on Social Security of the House Comm. on Ways and Means*, 96th Cong., 1st Sess. 78 (1979).

[10] Health and Human Services briefing paper, enclosure to letter from Patricia E. Dilley to Martha Derthick (Oct. 12, 1988).

political criteria can deprive an agency of effective leadership and advocacy during the crucial transition period between administrations, when new policies often are drafted.

Given the somewhat technical topic of *Agency Under Stress,* the book is surprisingly readable. Derthick's prose is rarely elegant, but it is clear and forthright, and the book's structure—as an inquiry into what went wrong, a kind of political whodunit—even injects a bit of drama into her exposition. Also noteworthy is her use of interviews with former agency officials and internal documents they supplied. These sources enable her to draw a convincing portrait of the psychological as well as political forces at work in the SSA. The book's readability occasionally comes at the price of specificity and detail. Some readers might appreciate, for instance, a more complete account of the actual features of SSI, or a fuller discussion of the legal theory underlying the SSA's policy of nonacquiescence. Other readers might also like to see the book take a more critical stance on some of the topics it discusses. Insofar as Derthick is concerned with process more than politics, she tends to withhold judgment on the substance of SSA policy, even when that policy has appeared to some as an example of "official lawlessness."[11] Derthick typically takes a more detached position, stating simply that the "legal merits of the SSA's position on nonacquiescence were open to debate" (p. 142).

To what extent can the kinds of problems described in *Agency Under Stress* be avoided in the future? Derthick notes at the outset of her book that, to a certain degree, administrative inefficiency and conflict are inherent to the American system of government (p. 4). On the one hand, efficiency and accountability would seem to dictate the consolidation of power over administrative agencies in the executive branch. On the other hand, the system of checks and balances justifies the involvement of Congress as appropriator of funds, lawmaker, and watchdog, while the courts serve as overseers. Conflict between these different interests, and consequently between the branches, ensures that administration of government programs will never be smooth in the American system. Derthick concludes:

> The summary answer, then, to how national administrative agencies fit into the American system is that they fit uneasily,

[11] Lewis, *Respect for Law?*, N.Y. Times, June 18, 1984, at 19, col. 1.

> under stress. . . . [M]uch in the U.S. constitutional tradition casts doubt on the legitimacy of whatever power they possess. . . . [B]y combining executive, legislative, and adjudicative functions, they flout the separation of powers. Yet they are also conspicuous victims of that separation, for they are often the focus of conflicting interaction among the three primary branches, each of which has a solid claim to a right to supervise them (p. 19).

Such problems, Derthick later notes, are the "price for the benefits of a system that is responsive to a pluralistic society" and "which values diffusion of governmental power" (p. 214).

Derthick argues, nevertheless, that the performance of administrative agencies can be improved. Her fundamental precept is that

> [i]t needs to be recognized that administrative considerations represent a legitimate, even central and urgent, claim upon the attention of all policymakers; that organizational capacities are a necessary and proper topic of reasoned inquiry, integral to policymaking; and that responsibility for nurturing those capacities—and using them intelligently—is borne by all the constitutional branches, not just by agency heads (p. 216).

More specifically, Derthick suggests that congressional staff agencies such as the General Accounting Office ("GAO") should take responsibility for producing, once every five or ten years, a comprehensive in-depth assessment of the performance and capacities of each major agency, with suggestions for improvement (pp. 217–20). Such reports, Derthick hopes, would increase policymakers' awareness of administrative issues in general and lead to more informed decisions.

Furthermore, she suggests that the GAO, the OMB, or the Congressional Budget Office should be asked to estimate the administrative costs and consequences of new legislative proposals, considering such issues as: how a program would be administered, and by whom; what kinds of new delivery systems and technologies it would require, and how much time would be needed to develop them; what sorts of burdens it would impose on federal, state, and local employees, as well as on the private sector; and what kinds of incentives and disincentives it would create both for beneficiaries and for government administrators (pp. 220–21). Finally, Derthick argues that as government becomes larger and its tasks more complex, policymakers will have to content themselves increasingly with piecemeal, incremental changes that are largely implemented by the states, rather than by the federal government:

> When programs involving a large volume of cases and many discretionary decisions at dispersed locations are being administered in a vast and socially heterogeneous country, centralization has its limits. . . . The center tends to lose control as volume and complexity increase in programs that necessarily lodge a great deal of discretion at the field level . . . (p. 222).

This last passage may sound like an endorsement of Reagan's New Federalism as an inevitable feature of modern government, but as Derthick shows in the case of disability review, even the Reagan administration was guilty of big-government hubris in trying to impose a massive, comprehensive change on a huge program in a very short period of time. Derthick is a pragmatist, not an ideologue. Unlike Reagan, she seeks not to dethrone the federal government, but only to point out some of the limits to its power.

Derthick's remedial proposals are relatively modest, but implementing them might improve greatly the performance of our administrative agencies. A still more effective proposal might be to require all new arrivals in Washington to read *Agency Under Stress*, for Derthick's cautionary tale of failed calculations and reasonable intentions gone wrong should awake both the visionary and the cynic to the need to pay attention to the nuts and bolts of administration.

—*A.W. Phinney*

ABORTION: THE CLASH OF ABSOLUTES. By *Laurence H. Tribe*. New York: W.W. Norton & Company, 1990. Pp. xvi, 259, notes, index. $19.95 cloth.

One need look no further than the morning paper to gauge the ubiquity of the abortion issue. The issue pervades public discourse, raising emotions and creating divisions. Pro-life, pro-choice, anti-abortion, pro-abortion: the wearisome and acrid polemic seems cast in stone.

Thus, any honest attempt to challenge the inevitability of permanent conflict on "the bitter and divisive public question of abortion 'policy'" and "to lay the groundwork for moving on" (p. 7) is welcome. When the invitation to soul-search is extended by a pre-eminent constitutional scholar whose academic and advocacy work has virtually defined one of the poles, the offer is both noteworthy and hopeful.

In *Abortion: the Clash of Absolutes*, Professor Laurence Tribe of Harvard Law School sets out to approach abortion anew (p. 3) and to "explore the legal framework in which the constitutional question of abortion rights must be decided" (p. 8). Avowing that it is not the book's goal "to 'prove' to anyone the correctness of any particular position in the abortion debate" (p. 8), he invites the reader on a journey from *Roe v. Wade*[1] to *Webster v. Reproductive Health Services*[2] in order "to consider whether *Roe* was rightly decided in the first place" (p. 26). His initial methodology: to examine "the history of the legal treatment of abortion in the United States," the "responses of other nations and cultures to the abortion question," and then "what the Constitution does or doesn't say about abortion" (p. 26).

In a brisk narrative spanning "two centuries of abortion in America" (p. 27), Tribe recounts the historical, cultural, and religious backdrop of the early restrictive statutes which outlawed abortion (pp. 27–34); "the metamorphosis of the abortion question into a matter of 'medical judgment'" (p. 34); the American Law Institute's ("ALI's") 1959 revision to its Model Penal Code which thereafter sanctioned abortion in cases of risk to the mother's health, likely birth defects, and rape or incest (p. 36); the thalidomide and rubella tragedies of the early 1960's, which sensitized physicians to the "quality of life" argument (p. 37); and the shift from a movement in the state legislatures to reform restrictive abortion laws to a crusade for the repeal of abortion restrictions altogether (pp. 42–49). The story is riveting. Tribe's admitted reliance, however, on a partisan brief supporting the Reproductive Health Services in *Webster*[3] as the "point of departure" for much of the historical data in the chapter (p. 244 n.1) is somewhat regrettable in a book seeking common ground.

---

[1] 410 U.S. 113 (1973) (ruling that abortion is a woman's "fundamental" right embedded in the previously recognized right of "privacy," abridgeable by government only when demonstrably necessary to achieve a "compelling" objective (p. 11)).

[2] 109 S. Ct. 3040 (1989) (plurality opinion) (reducing a woman's right to abortion to a "liberty interest" more easily abridgeable, but only by government regulation which does not impose an "undue burden" on a woman's abortion decision (p. 23)).

[3] Brief of 281 American Historians as Amici Curiae Supporting Appellees, Webster v. Reproductive Health Services, 109 S. Ct. 3040 (1989) (No. 85-605).

This section also occasions Tribe's first dispute with Professor Mary Ann Glendon of Harvard Law School. In *Abortion and Divorce in Western Law*,[4] Glendon argues that *Roe* "interrupted an evolutionary process within state legislatures," and that, if *Roe* were to be repealed, "abortion would continue to be freely available . . . during the first trimester of pregnancy" (p. 49). Tribe disagrees, arguing that the movement for the repeal of laws restricting abortion was not as great as Glendon suggests, and that a repeal of *Roe* would result in "a meaningful decline in the availability of abortion services . . ." (p. 51).

Tribe's ensuing survey of the treatment of the abortion issue throughout the world illustrates the universality of abortion throughout time and place, regardless of prohibition (p. 52). According to Tribe, history seems to instruct us that oppressive regimes uniformly outlaw abortions: Stalin's Soviet Union (p. 56), communist Romania, beginning in 1966 (p. 57), and Hitler's Germany, the last of which, according to Tribe, best exemplifies the potential evils which arise with government control over abortion (p. 59). In an interesting rhetorical inversion of the characterization of abortion by pro-life groups as the American Holocaust, Tribe associates Nazi Germany with an emphasis on the duty to have large families, the shutting down of family planning clinics, and the control of contraception (p. 59).

Furthermore, he counsels, "ideas about abortion and attitudes about its availability are highly culture-specific" (p. 76). Therefore, any attempt to re-orient the "individual rights" emphasis of the abortion debate in America (p. 52) towards a European-style "compassion with affirmation of life" framework (p. 72), as advocated by Glendon in her book, will not work. Our confidence in the rule of law, our insistence on enforceable norms, our "uniquely American ideology of individual worth that has led us to a largely rights-based legal system" (p. 74), would lead us to reject any empty "life-affirming" normative statement of principle (pp. 73–74).

The extent to which Tribe is willing to carry this libertarianism—at least in the abortion context—is evident in the following chapter's discourse on constitutional analysis. Having acknowledged that the absolutes clashing in the abortion controversy are life and liberty (p. 3), he subordinates the former to the latter, extolling the "long tradition of asking *first* whether the

---

[4] M.A. Glendon, Abortion and Divorce in Western Law (1987).

right asserted is a fundamental liberty," and only then considering the compelling reasons, such as the life interest of a fetus, which might justify its abridgement (p. 96). Justice Antonin Scalia, writing in *Michael H. v. Gerald D.*,[5] questioned in another context this "[s]trange procedure of looking at the act which is assertedly the subject of a liberty interest in isolation from its effect *upon other people*—rather like inquiring whether there is a liberty interest in firing a gun where the case at hand happens to involve its discharge into another person's body" (p. 96).[6] Tribe rejoins that Scalia's holistic approach to examining rights would merely result in defining that right in terms of the state's interest, which would ultimately "do violence to all our rights" (p. 97). It would, furthermore, "simply deny a woman, in the first instance, even a hard look at the reasons why her ability to choose abortion was being restricted . . ." (p. 97).

Similar libertarian reasoning was employed in the case of *Lochner v. New York*,[7] and its progeny, which imputed the substantive liberty interests of contract and property to the due process clause of the fourteenth amendment. The isolated evaluation of those "fundamental" economic liberties permitted the invalidation of "child labor laws, limitations on the hours of work, laws regulating labor-management relations . . . [, and] all regulations of the employment relationship except those the Court found to be directly in aid of the *public* health and welfare . . ." (p. 85). While acknowledging the folly of an emancipation that in practice amounted to what one commentator called "the legal right to starve" (p. 85),[8] Tribe nonetheless defends *Lochner*'s "liberty first" rationale, finding fault only with its "misguided understanding of what 'liberty' required" (p. 86). He argues, as did the Court in *Lochner*, that the liberty clause of the fourteenth amendment guarantees "substantive protections of individual rights from intrusion by the government" (pp. 83–84). In the case of abortion, the right is that of "privacy": a concept articulated by Justice Louis Brandeis as "the right to be let alone" (p. 92)[9] that had expanded, via *Griswold v.*

[5] 109 S. Ct. 2333 (plurality opinion).
[6] *Id.* at 2341 n.4 (emphasis in original).
[7] 198 U.S. 45 (1905) (invalidating New York's 60-hour limit on a bakery employee's workweek on the grounds that it interfered with a worker's liberty to contract and with the employer's freedom to use his property in traditional ways (pp. 84–85)).
[8] *The Legal Right to Starve*, 34 New Republic 254 (May 2, 1923).
[9] Olmstead v. United States, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting).

Connecticut[10] and its progeny, into a fundamental "right to engage in sexual intercourse without having a child" (p. 94).

Tribe links abortion rights to contraception rights in several passages and deftly uses the fear of losing the latter as an incentive for protecting the former. For instance, he questions rhetorically, if *Roe* could be overturned, and "if women have no significant 'liberty' at stake in the abortion context, how can they possess a fundamental liberty to use birth control?" (p. 95). He warns that countenancing the argument that only enumerated rights legitimately bind the states through the fourteenth amendment, an argument frequently invoked to disparage *Roe*, could create a regime in which the federal and state governments could abolish birth control as well as abortion (p. 90). He warns that acknowledging the fetus as a person from the moment of conception would likely result in the reversal "in significant part of the Supreme Court's long line of contraception decisions" (p. 122), since many forms of contraception are abortifacient, and the liberty to use such methods could be outweighed by the constitutional protection of life (p. 122). Tribe likely is correct in his presumptions. By piquing our self-interest in non-procreative sex, however, he seems to entice us into evaluating the constitutional and philosophical merits of specific propositions with an instinct not noted for its reasonableness.

Tribe identifies a woman's privacy interest as "the right not to remain pregnant" (p. 98), expressed alternatively as the "liberty not to be made unwillingly into a mother" (p. 98), and (most familiarly) the right to decide whether or not to choose to terminate a pregnancy (p. 141). This liberty can be abridged only in the case of a compelling governmental interest, and, until technology permits otherwise, this right subsumes the analytically distinct "right to destroy one's fetus" (p. 98). Killing the fetus is simply the unavoidable "outcome" of abortion, an outcome not necessarily protected by *Roe* (p. 115).

Tribe argues that the consequence of denying this liberty interest in decisions about reproduction is a step onto the slippery slope of government tyranny and toward mandated abortions and sterilizations (p. 111). Tribe's argument is that if the state can control one reproductive decision, it can control them all. This is true, of course, only insofar as liberty trumps life.

---

[10] 381 U.S. 479 (1965) (interpreting the liberty clause to protect the right of a married couple to decide whether to use contraceptives (p. 93)).

Were the life interest paramount, then there would be no such threat of either forced abortions or sterilizations. Naturally, abortion on demand would also be impeded in such a regime since in a "life first" analysis, the "right to kill a fetus . . . wouldn't be fundamental by anyone's definition" (p. 97).

Tribe then inquires into the relevance of the fetus's personhood. Acknowledging that "the scientific 'disproof' of separate embryonic personhood . . . cannot succeed completely" (p. 119), he observes that "the Constitution uses the word 'person' in a way that would not really make sense if fetuses were thought to be included" (p. 120). Regardless, "[a] woman denied the right to decide whether or not to end a pregnancy is not merely being asked to refrain from killing another person but being asked to make an affirmative sacrifice . . . in order to save that person" as well (p. 130). Noting that such sacrifice is not traditionally required, he deduces that a law prohibiting a woman from freeing herself of the fetus inside her would "work a harsh discrimination against women *even if fetuses count as persons*" (p. 131). To impose virtue on any person demeans that person's individual worth, and "there should be no 'woman's exception' to our traditional regard for individualism and autonomy" (p. 135).

Tribe distinguishes his position from that of those who pretend that there is only one party—the woman—in the picture. He grants that "as pregnancy progresses the fetus's value becomes ever harder to deny . . . [and] the moral picture reveals two beings" (p. 138). He thus endorses the wisdom of *Roe*'s "viability" approach, which reflects "the widely shared sense that we should erase neither the fetus *nor* the woman from the picture our law presents" (p. 138). In practice, however, *Roe*'s dual perspective proved illusory since, as Tribe notes earlier in the book, the Court struck down every state restriction on abortion it considered for sixteen years after *Roe* until finally upholding one in *Webster* (p. 24).

Tribe resumes with narrative history in the ensuing two chapters, which review developments from *Roe* to *Webster* and beyond. He recounts the initial pro-choice and pro-life reactions to *Roe*, with the resulting ascent of a largely religious "New Right" (pp. 139–50); the split in the pro-life movement between those who wished to return the issue to the states and those who would prohibit all or most abortions (p. 148); the 1976 election with its conservative Republican "right to life" plank and the "less adamant" Democratic position opposing a constitutional amendment to overturn *Roe* (p. 149); the emergence of Rep. Henry Hyde (R-Ill.) and the ban on Medicaid funding for abortion with which he is closely associated (pp. 151–59); the election of Ronald Reagan in 1980 (p. 161) and the subsequent failed legislative attempts to reverse *Roe* (pp. 161–65); Reagan's revolutionizing of the judiciary (pp. 167–70); the pre-*Webster* political posturing and the crystallization of the "Who decides?" formulation of the abortion issue (pp. 172–76); and, finally, the post-*Webster* pro-choice backlash and the Republican waffling which greeted the return of the abortion issue to the federal and state legislatures (pp. 177–92). Tribe reports "that if the pro-choice movement is to maintain its momentum, it cannot let the pro-life side shift the debate to *why* a woman wants any given abortion" since its "current popular appeal clearly depends on keeping the question focused on *who* will make the decision" (p. 193).

Tribe is critical of most of the currently discussed compromises. Spousal consent laws "would simply transfer power from the woman, who has decided on an abortion, to the man, who has decided to stop her" (p. 198); parental consent laws raise the specter of parentally compelled abortions (p. 199), "and are likely in practice to achieve little while causing great grief" (p. 203); mandatory waiting periods "will act as an absolute obstacle for at least some women who might otherwise obtain legal abortions" (p. 204); limiting the reasons for which abortion will be allowed invades privacy (p. 205); restrictions on government funding is an "immoral" denial of the right to choose to terminate a pregnancy to poor women (p. 207); regulating abortion clinics adds significant costs to the provision of abortion services, "driving some abortion clinics out of business or greatly raising the price of an abortion for a woman seeking one . . ." (p. 207); earlier cutoff dates impose arbitrary limits and satisfy neither those who identify personhood at the moment of conception nor those who believe in a woman's right to self-determination (p. 208). For these reasons, Tribe argues, all of these recommendations should be rejected.

Tribe's compromises include affordable postnatal health care, mandatory maternity and paternity leaves, good child care, and flexible work time arrangements (p. 211); "sex education and the wide availability of birth control" (p. 212); and advanced contraceptive technologies and abortifacients including RU-486 (pp. 213–15). These recommendations will lead to a world of "only wanted pregnancies," the desire for which is "the common ground on which we all can stand" (p. 228).

Tribe closes by transcending the absolutes of a fetus's right to life and the woman's right to liberty. "Neither 'absolute' is really that," he suggests (p. 230). Future technology might separate the two analytically distinct abortion issues of the imposition on a woman's liberty and the destruction of the fetus for which she is responsible (p. 222). If "one's pro-choice views rest on a sense that a woman should be allowed to prevent her fetus from becoming a child even if it could become one without the woman having to undergo a prolonged pregnancy, then perhaps, in a technologically transformed world, those views would have to yield to the claim to life of all but the most undeveloped fetus" (p. 225). Hence, a woman's liberty in the future might not be so absolute as to empower her to kill her fetus without necessity. As for the present, Tribe suggests that rather than from any absolute reverence for life or commitment to the protection of unborn children, "the feeling that abortion should be blocked by government may grow, at least in part, out of a reflexive willingness to enforce traditional sex roles upon women and to impose upon them an unequal and harsh sexual morality" (p. 237).

Thus, we arrive at the denouement. The abortion polemic is not really a clash of absolutes, one between life and liberty. For the present, at least until science progresses, it is a clash between a contingent and an absolute: between repression and liberty.

Tribe's book has much to recommend it. It is thorough and extremely well organized. He raises nearly all of the issues, and deals squarely with most of the arguments. He presents his abortion adversaries at their best—Glendon, Scalia, Robert Bork, Chief Justice William Rehnquist, John Hart Ely, John C. Willke, and Randall Terry—and confronts them all. With few exceptions, he erects no straw men to beat down in their stead. The arguments are clever, but what impresses the reader most is that they are the skillfully crafted offerings of a passionate advocate who is acutely aware that "[i]n a democracy, voting and persuasion are all we have" (p. 240).

Perhaps it is this latter fact that disappoints the most, as well as impresses. Compromise often requires the work of a dispassionate scholar, as well as that of a passionate advocate. Yet one realizes throughout that Tribe is unable to set aside his partisan advocacy of abortion rights. For instance, he castigates the Human Life Statute sponsored by Senator Jesse Helms (R.-N.C.) in the early 1980's, which "would have defined 'person' to include an embryo from the moment of conception[,]" as a "fatally flawed idea . . . in fundamental conflict with our very structure of government" (p. 162). The bill's disregard for the Framers's clear intention that constitutional doctrine be legislatively overturned only by constitutional amendment, and not by a "mere majority vote in Congress" (p. 162), evidently does not apply, in Tribe's mind, to the unsuccessful Freedom of Choice Act of 1989, which would have preserved "*Roe v. Wade* through federal statutory law, by prohibiting the states from imposing restrictions on [most] abortions . . ." (p. 191). Rather than criticizing the Freedom of Choice Act, Tribe instead offers his constitutional rationale for the statutory overturning of *Webster* that "leaves open a continuing possibility of future involvement by Congress in any political resolution of the abortion issue" (p. 192).

Tribe's philosophical ambiguities are also worrisome, as when, after trumpeting rugged individualism and belittling Glendon's suggestion of importing the Europeans' communitarian approach to the abortion problem, he nonetheless appropriates as his own solution the social welfare net that the communitarian approach engenders. The shift from a libertarian conception of rights to a communitarian notion of responsibilities goes unnoted and unexplained. Also unillumined is the process by which focusing the abortion debate primarily on a woman's autonomous liberty, in isolation from the results that follow from the exercise of this liberty, furthers the causes of "humane concern," "compassionate impulse" (p. 211), "genuine reverance for life," and "true respect for all humanity" (p. 196). Unless, as with Adam Smith's capitalist,[11] each autonomous self-interested actor is led by an "invisible hand" that promotes unintended ends, it is unclear how each woman, by freely promoting her own interest, promotes humanitarianism and compassion for others better than if she were trying to promote them directly. The metaphysics necessary to make such a phenomenon plausible are not set forth for examination.

---

[11] A. SMITH, THE WEALTH OF NATIONS (1982).

Abortion: The Clash of Absolutes is an informative and interesting book. Perhaps understandably, however, Tribe takes sides, and ultimately he is not able to "approach abortion anew." It is nonetheless a book worth reading, for the insights it stimulates as well as for the ones it offers.

—*Maximilian B. Torres*