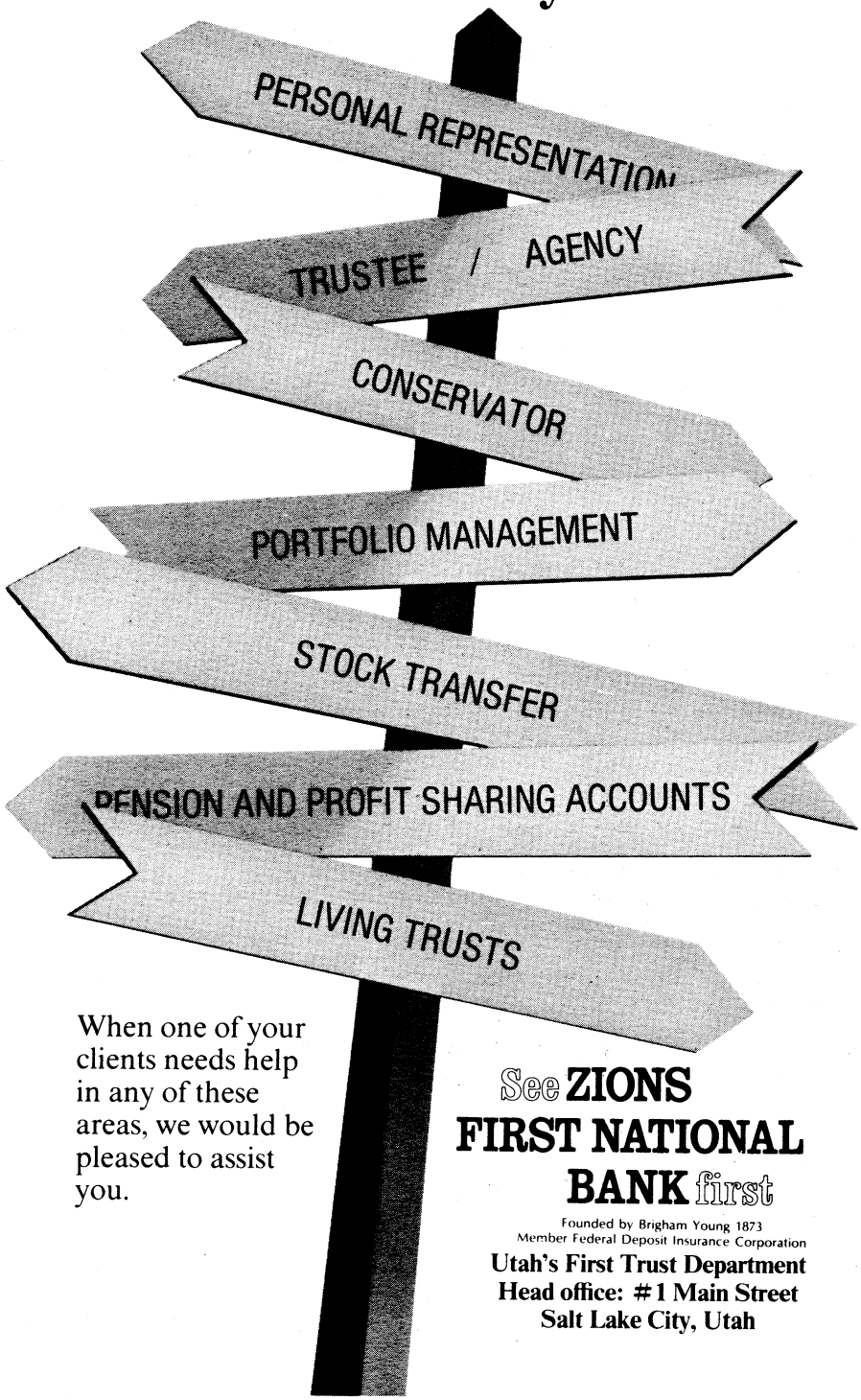
VOLUME 1977 NUMBER 4

UTAH LAW REVIEW

UNIVERSITY OF UTAH COLLEGE OF LAW

Wherever you're going... we can serve you.



When one of your clients needs help in any of these areas, we would be pleased to assist you.

See **ZIONS FIRST NATIONAL BANK** first

Founded by Brigham Young 1873
Member Federal Deposit Insurance Corporation

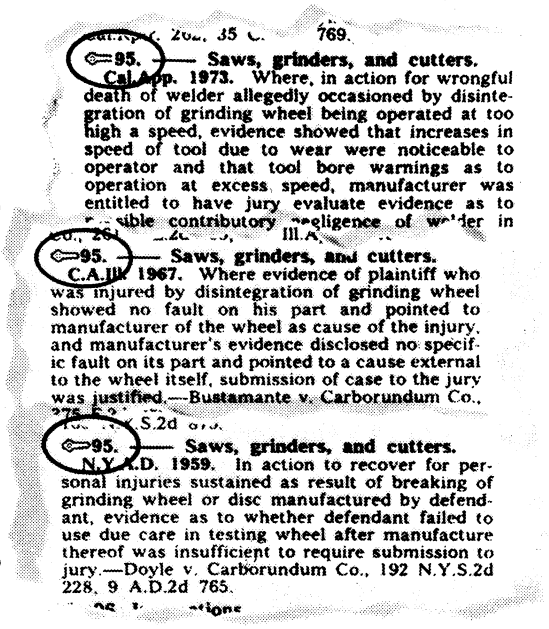
Utah's First Trust Department
Head office: #1 Main Street
Salt Lake City, Utah

One Key Number gives me all cases in point

in California...

in Illinois...

in New York...



769.

⇐95. —— **Saws, grinders, and cutters.**

Cal.App. 1973. Where, in action for wrongful death of welder allegedly occasioned by disintegration of grinding wheel being operated at too high a speed, evidence showed that increases in speed of tool due to wear were noticeable to operator and that tool bore warnings as to operation at excess speed, manufacturer was entitled to have jury evaluate evidence as to ...sible contributory negligence of w'ler in

...Ill.A.

⇐95. —— **Saws, grinders, and cutters.**

C.A.Ill. 1967. Where evidence of plaintiff who was injured by disintegration of grinding wheel showed no fault on his part and pointed to manufacturer of the wheel as cause of the injury, and manufacturer's evidence disclosed no specific fault on its part and pointed to a cause external to the wheel itself, submission of case to the jury was justified.—Bustamante v. Carborundum Co., ...S.2d ...

⇐95. —— **Saws, grinders, and cutters.**

N.Y.A.D. 1959. In action to recover for personal injuries sustained as result of breaking of grinding wheel or disc manufactured by defendant, evidence as to whether defendant failed to use due care in testing wheel after manufacture thereof was insufficient to require submission to jury.—Doyle v. Carborundum Co., 192 N.Y.S.2d 228, 9 A.D.2d 765.

...ions

in any jurisdiction in the United States

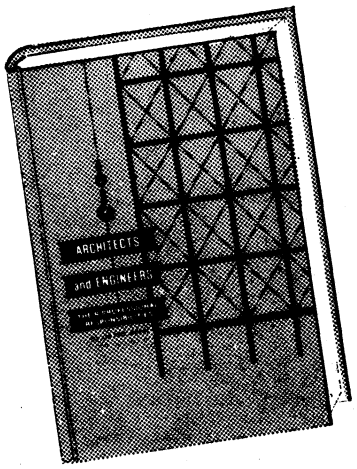
With West's exclusive Key Number System, I can check my local Key Number Digest and instantly have access to all reported cases on that point. And this same Key Number will work in any other state digest, in any jurisdiction in the United States. This gives you instant access to all cases of all jurisdictions.

Save time and work. Start your search with West's Key Number System. To get full details, write today.

WEST'S KEY NUMBER SYSTEM
—the exclusive research system for lawyers.

WEST PUBLISHING COMPANY

Wallace A. "Buck" McLean
4135 Cumberland Road
Salt Lake City, UT 84117
Phone: 801/277-8072



ARCHITECTS and ENGINEERS

THEIR PROFESSIONAL RESPONSIBILITIES

By James Acret, J.D.,

member State Bar of California

This book was written for lawyers, engineers, architects and insurance claims departments. It is a complete work on the subjects of responsibilities, liabilities and malpractice of architects/engineers. It reviews the various insurance coverages available and recommends specific malpractice avoidance techniques for engineers and architects to adopt in their respective practices.

For the lawyer, there is a comprehensive review of the nationwide authorities on these subjects along with information on how to represent engineers and architects in malpractice cases.

- ☐ Please send me ARCHITECTS and ENGINEERS, their professional responsibilities, and send the future supplements as published.
- ☐ In 15 days I will send $11.00 and $10.00 per month for 3 months.
- ☐ Payment of $40.00 enclosed. Shepard's to pay postage and handling.

NAME .....

FIRM .....

STREET .....

CITY.....STATE.....ZIP.....

Subject to acceptance in Colorado Springs, Colorado

1 Volume, 449 Pages, 6x9, Hard Cover $40.00
Plus Postage and Handling . . . . . $ 1.00
Annual Supplementation

Shepard's, Inc. of Colorado Springs
P. O. BOX 1235, COLORADO SPRINGS, COLORADO 80901

UTAH LAW REVIEW

VOLUME 1977 NUMBER 4

EDITORIAL OFFICES: The *Utah Law Review* is published at the University of Utah College of Law, Salt Lake City, Utah 84112 by the Utah Law Review Society. Second-class postage paid at Salt Lake City, Utah, and at additional mailing offices. All communications should be sent to the editorial offices.

SUBSCRIPTIONS: The *Utah Law Review* is published quarterly. The subscription rate is $15 for one year, $28 for two years, and $40 for three years. Single issues may be obtained for $4.00 per issue and reprints of articles and notes for $2 per copy. If the subscriber wishes his subscription discontinued at its expiration, notice should be sent; otherwise, it is assumed that continuation is desired.

Printed in the United States of America
by Darby Printing Company

UNIVERSITY OF UTAH

COLLEGE OF LAW FACULTY

AARON, RICHARD I. (1966), Professor of Law. A.B., 1959, Harvard University; J.D., 1962, University of Wisconsin.

ANDERSON, JERRY R. (1963), Professor of Law. B.S.L., 1955, J.D., 1957, University of Utah.

BAYLES, J. WENDELL (1973), Professor of Law. B.S., 1964, J.D., 1965, University of Utah; LL.M. (Taxation), 1966, New York University.

BENNETT, WALLACE R. (1953), Professor of Law. B.A., 1949, M.S., 1950, J.D., 1953, University of Utah; LL.M., 1957, Harvard University; LL.M. (Trade Regulations), 1963, New York University.

BOYCE, RONALD N. (1966), Professor of Law. B.S.L., 1955, J.D., 1957, University of Utah.

DYER, BOYD K. (1971), Professor of Law. B.A., 1962, Stanford University; J.D., 1968, Harvard University.

EMERY, ALFRED C. (1947), Professor of Law. B.S., 1940, J.D., 1947, University of Utah.

FIRMAGE, EDWIN B. (1966), Professor of Law. B.S., 1960, M.S., 1962, Brigham Young University; J.D., 1963, LL.M., 1964, S.J.D., 1964, University of Chicago.

FLYNN, JOHN JOSEPH (1963), Professor of Law. B.S., 1958, Boston College; LL.B., 1961, Georgetown University; S.J.D., 1967, University of Michigan.

FORDHAM, JEFFERSON B. (1970), Professor of Law, former Dean of the University of Pennsylvania College of Law. A.B., 1926, M.A., 1929, J.D., 1929, University of North Carolina; S.J.D., 1930, Yale University; LL.D., 1953, University of North Carolina.

FRANKEL, LIONEL H. (1966), Professor of Law. B.A., 1953, Ursinus College; J.D., 1956, Yale University; LL.M., 1962, New York University.

LOCKHART, WILLIAM J. (1964), Professor of Law. B.A., 1955, J.D., 1961, University of Minnesota.

OBERER, WALTER E. (1975), Dean of the College of Law and Professor of Law. B.A., 1942, Ohio Wesleyan; LL.B., 1948, Harvard University.

PETERSON, ROBERT (1976), Associate Professor of Law. J.D., 1971, University of Utah.

SCHMID, ROBERT L. (1954), Professor of Law. J.D., 1954, University of Utah; M.L.L., 1956, University of Washington.

STRACHAN, KRISTINE (1973), Associate Professor of Law. B.A., 1965, University of Southern California; J.D., 1968, University of California.

STRONG, KLINE D. (1964), Adjunct Professor of Law. A.B., 1951, Brigham Young University; M.B.A., 1952, Northwestern University; LL.B., 1955, University of Colorado; Ph.D., 1970, University of Utah.

SWENSON, ROBERT W. (1953), Professor of Law. B.S.L., 1940, J.D., 1942, University of Minnesota.

THODE, E. WAYNE (1965), Professor of Law. B.S., 1943, University of Illinois; LL.B., 1950, University of Texas; S.J.D., 1964, Harvard University.

THURMAN, SAMUEL D. (1962), Professor of Law, former Dean of the University of Utah College of Law. A.B., 1935, University of Utah; J.D., 1939, Stanford University.

VAN ALSTYNE, ARVO (1966), Professor of Law. B.A., 1943, J.D., 1948, Yale University.

WILKINS, RUTH W. (1966), Professor of Law and Law Librarian. J.D., 1953, University of Utah.

ZIMMER, BRUCE I. (1975), Associate Dean of the College of Law. A.B., 1965, Columbia College; J.D., 1968, New York University.

ZIMMERMAN, MICHAEL D. (1976), Associate Professor of Law. B.S., 1967, University of Utah; J.D., 1969, University of Utah.

UTAH LAW REVIEW

Member, National Conference of Law Reviews

VOLUME 1977 NUMBER 4

UTAH LAW REVIEW

VOLUME 1977 NUMBER 4

TABLE OF CONTENTS

Fairly Due Process: Minimum Protection Recognized but not Applied in *Mathews v. Eldridge*

*William H. Lawrence**

The problem is to set the margin of tolerable error, given the ills of too much procedure.[1]

FRANK C. NEWMAN

Over a decade and a half have passed since Professor Newman succinctly identified the major unanswered concern with the application of procedural due process. Since then, and particularly in the 1970's, due process has taken on radical new contours,[2] although the changes have not all been in one direction. In a 1975 lecture, Judge Henry J. Friendly stated: "[W]e have witnessed a greater expansion of procedural due process in the past five years than in the entire period since ratification of the Constitution."[3] The opinions in several subsequent cases suggest, however, that the Supreme Court is now attempting to retreat from its more expansive positions on procedural due process.[4] A number of cases decided in 1976 tended to restrict the breadth of liberty and property interests encompassed within the meaning of the due process clauses.[5] Another

---

* Associate Professor of Law, University of Toledo College of Law. The author wishes to acknowledge the capable research assistance of Oatfield Whitney III, second-year law student, University of Toledo.

1. Newman, *The Process of Prescribing "Due Process,"* 49 CAL. L. REV. 215, 228 (1961) (emphasis in original).

2. Previously, procedural due process protections were extended only to recognized rights; mere privileges could be summarily withdrawn. A leading case was Bailey v. Richardson, 182 F.2d 46 (D.C. Cir. 1950), *aff'd per curiam by an equally divided court*, 341 U.S. 918 (1951). The Supreme Court began dissolving the privilege doctrine in the mid-1960's. *See* Reich, *Individual Rights and Social Welfare: The Emerging Legal Issues*, 74 YALE L.J. 1245 (1965); Reich, *The New Property*, 73 YALE L.J. 733 (1964); Van Alstyne, *The Demise of the Right-Privilege Distinction in Constitutional Law*, 81 HARV. L. REV. 1439 (1968). By 1971 the Court had totally rejected the concept. Graham v. Richardson, 403 U.S. 365, 374 (1971).

In 1970, the Supreme Court decided the landmark case of Goldberg v. Kelly, 397 U.S. 254 (1970). Since then the Court has decided thirteen major procedural due process cases. *See* K. DAVIS, ADMINISTRATIVE LAW OF THE SEVENTIES §§ 7.00 to 7.00-11 (1976), 7.00-1-1 (Cum. Supp. 1977).

3. Friendly, *"Some Kind of Hearing,"* 123 U. PA. L. REV. 1267, 1273 (1975).

4. For a discussion of the development of procedural due process law during this period, see Lawrence, *A Restatement of the Roth-Fuentes Analysis of Procedural Due Process*, 11 GA. L. REV. 477 (1977).

5. Bishop v. Wood, 426 U.S. 341 (1976) (discharge of police officer; held no liberty interest because the asserted reasons for discharge were communicated to the police officer and not to the public, and no property interest because the ordinance was interpreted not to confer tenure; 5-4 decision); Paul v. Davis, 424 U.S. 693 (1976) (distribution of picture as active shoplifter; held reputation alone does not invoke due process protection; 5-3 opinion);

opinion of the same year, *Mathews v. Eldridge*,[6] is the Court's latest statement concerning procedural requisites when due process protections apply. In different respects, this opinion is both expansive and restrictive.

The *Eldridge* case includes a general framework of analysis which, more than any other procedural due process opinion, points the proper direction to follow in resolving the enigma posed by Professor Newman. Unfortunately, the Court's application of that analysis serves to deny success in the endeavor. The *Eldridge* opinion represents an expanded role for the interests of the individual in theory, but not in practice. Having envisioned the means by which to narrow it with greater precision, the Court proceeds to set a margin of error that is too broad.

This critique on *Mathews v. Eldridge* is further developed through this article. Following a brief explanation of the case, the Court's holding as it relates to its positive framework of analysis is explained. The article next considers the inadequacies of the application of the *Eldridge* analysis. Suggestions on ways to improve the application of procedural due process so as to realize the benefits of the Court's new framework of analysis comprise the third and final section.

I. The Positive *Eldridge* Framework of Analysis

A. *Case Background*

The Social Security Act[7] provides for money benefits to workers who are completely disabled.[8] After establishing initial entitlement, the worker then bears the burden of proving continued eligibility. He must demonstrate by "medically acceptable clinical and laboratory diagnostic techniques"[9] that he is unable "to engage in any substantial gainful activity by reason of any medically determinable

---

Meachum v. Fano, 427 U.S. 215 (1976) (transfer of prisoner to prison with less favorable conditions; held no liberty interest absent state law or practices making transfer contingent upon misconduct or other events; 6-3 opinion); Montanye v. Haymes, 427 U.S. 236 (1976) (same as Meachum case; 6-3 opinion); Kelley v. Johnson, 425 U.S. 238 (1976) (haircut regulations for policemen; held no protected interest; 6-2 opinion).

6. 424 U.S. 319 (1976). Professor Davis has written that this opinion "may penetrate further than any previous Supreme Court opinion on this general subject." K. Davis, *supra* note 2, § 7.00-1-1, at 85 (Cum. Supp. 1977).

7. 42 U.S.C. §§ 301-1396 (1970 & Supp. V 1975).

8. The physical and mental disability insurance program is covered by Title II of the Social Security Act, 42 U.S.C. §§ 405-432 (1970 & Supp. V 1975). Although the Social Security Act provides for other types of benefits, this article, like the *Eldridge* opinion, deals specifically only with the title II program.

9. *Id.* § 423(d)(3) (1970).

physical or mental impairment."[10]

Continuing eligibility assessments[11] are made by a state agency using information supplied by the worker and those persons giving him medical treatment and, in some cases, by independent medical examinations. A final determination by the state agency to terminate benefits is reviewed by the Social Security Administration (SSA). When the SSA concurs with the state agency, it sends written notification to the recipient and benefits are terminated effective two months following the month in which the worker's recovery is determined to have occurred. The worker is granted the right to a *de novo* reconsideration by the state agency; an adverse decision entitles the worker to an evidentiary hearing before an administrative law judge.[12] Discretionary review by the SSA Appeals Council and judicial review are the procedural avenues for appeal of decisions against the worker. The recipient is statutorily entitled to retroactive payments for any benefits withheld during a period in which the disability is ultimately found to have been continuing.[13]

The respondent in *Mathews v. Eldridge* was notified by the SSA of its acceptance of the state agency decision that he was no longer disabled. He was informed that his benefits would end after that month and that he could request state agency reconsideration of his case. Rather than seek reconsideration, the respondent filed suit in federal district court challenging the administrative procedures as being violative of due process.[14] The district court agreed

---

10. *Id.* § 423(d)(1)(A) (1970). The impairment must have lasted or "be expected to last for a continuous period of not less than 12 months." *Id.* The statute also provides that

> an individual . . . shall be determined to be under a disability only if his physical or mental impairment or impairments are of such severity that he is not only unable to do his previous work but cannot, considering his age, education, and work experience, engage in any other kind of substantial gainful work which exists in the national economy, regardless of whether such work exists in the immediate area in which he lives, or whether a specific job vacancy exists for him, or whether he would be hired if he applied for work.

Id. § 423(d)(2)(A).

11. Agency procedures for initial approval and for termination of benefits are included in 20 C.F.R. §§ 404.901 to .990 (1977).

12. 42 U.S.C. § 405(b) (Supp. V 1975); *id.* § 421(d) (1970); 20 C.F.R. § 404.917 (1977). For the views of an administrative law judge on the appeals process, and particularly the importance of representation by counsel, see Smith, *Social Security Appeals in Disability Cases*, 28 AD. L. REV. 13 (1976). The views of a practicing attorney are provided in Bloomfield, *Disability Claims Under the Social Security Act: A Practitioner's Guide to Administrative Procedures*, 6 CAP. U.L. REV. 371 (1977).

13. 42 U.S.C. § 404(a)(2) (1970); 20 C.F.R. §§ 404.501, .503, .504 (1977).

14. A previous procedural due process challenge on the termination of social security disability benefits reached the Supreme Court in Richardson v. Wright, 405 U.S. 208 (1972) (vacated as moot and remanded). In that case, however, the Court did not reach the constitutional issue. For a discussion of the case, see Meyerhoff & Mishkin, *Application of Goldberg*

with his position, holding that an evidentiary hearing prior to the termination of disability benefits was constitutionally required.[15] Following an affirmance by the court of appeals,[16] the Supreme Court reversed in a 6-2 decision.[17]

B. Tilting the Balance

The Court's analysis of the due process question begins with a summary of the factors identified in previous court opinions as the requisites of due process analysis. Justice Powell, writing for the majority, concludes that the Court must consider three factors:

> First, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government's interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.[18]

Since these criteria are derived from prior cases, and since each is discussed in relation to Eldridge's case, Justice Powell's opinion appears, at first glance, to simply measure the specific procedures for termination of social security disability benefits against a consistent framework of analysis of procedural due process.[19] Actually, the opinion makes a significant departure from the Court's previous approach.

The departure relates to the Court's use of the information generated through consideration of the three factors. Previous opin-

---

v. Kelly Hearing Requirements to Termination of Social Security Benefits, 26 STAN. L. REV. 549, 554-59 (1974).

15. Eldridge v. Weinberger, 361 F. Supp. 520 (W.D. Va. 1973).

16. Eldridge v. Weinberger, 493 F.2d 1230 (4th Cir. 1974).

17. Mathews v. Eldridge, 424 U.S. 319 (1976).

18. *Id.* at 335.

19. The listing of the three factors is clearly the most cited passage of the opinion. A large number of subsequent court opinions have quoted the precise language. Ong v. Tovey, 552 F.2d 305, 307 (9th Cir. 1977); Basel v. Knebel, 551 F.2d 395, 398 (D.C. Cir. 1977); Zurak v. Regan, 550 F.2d 86, 93 (2d Cir. 1977); Kennedy v. Robb, 547 F.2d 408, 414 (8th Cir. 1976); McGrath v. Weinberger, 541 F.2d 249, 253 (10th Cir. 1976); Johnson v. Mathews, 539 F.2d 1111, 1120 (8th Cir. 1976); Bignall v. North Idaho College, 538 F.2d 243 (9th Cir. 1976); Stretten v. Wadsworth Veterans Hosp., 537 F.2d 361, 367 (9th Cir. 1976); Betts v. Tom, 431 F. Supp. 1369 (D. Hawaii 1977); Aaron Ferer & Sons Co. v. Berman, 431 F. Supp. 847, 852 (D. Neb. 1977); Montrym v. Panora, 429 F. Supp. 393, 398 (D. Mass. 1977); Kickey v. New Castle County, 428 F. Supp. 606, 610 (D. Del. 1977); Diello v. City of Wilmington, 426 F. Supp. 1272, 1289 (D. Del. 1976); Graves v. Meystrik, 425 F. Supp. 40, 48-49 (E.D. Mo. 1977); Stolberg v. Caldwell, 423 F. Supp. 1295, 1300 (D. Conn. 1976); Smith v. Webb, 420 F. Supp. 600, 605 (E.D. Pa. 1976); Feinberg v. Federal Deposit Ins. Corp., 420 F. Supp. 109, 120-21 (D.D.C. 1976); Koger v. Guarino, 412 F. Supp. 1375, 1387 (E.D. Pa. 1976).

ions dictated the need in each case to balance the interest of the individual against the interest of the government so as to accommodate them both somehow in ascertaining the procedural safeguards required by the Constitution. Although the *Eldridge* opinion is also written in terms of considering and analyzing these interests,[20] the Court indicates, for the first time, that the scales are not equally balanced. As discussed below, the majority opinion evidences an intent to consider the interest of the individual as of paramount importance.[21]

The Court adopted the balancing test in the landmark opinion of *Goldberg v. Kelly*.[22] "The extent to which procedural due process must be afforded the recipient is influenced by the extent to which he may be 'condemned to suffer grievous loss,' . . . and depends upon whether the recipient's interest in avoiding that loss outweighs the governmental interest in summary adjudication."[23] In subsequent opinions, the Supreme Court purported to balance private and governmental interests to ascertain the timing and form of a hearing necessary to satisfy the requirements of procedural due process in particular situations.[24] The balancing approach, however, has not proven to be particularly successful. The specific requirements for a presuspension hearing, together with an explanation of the rationales underlying those requirements, were articulated in

---

20. At one point, the Court even spoke in terms of "striking the appropriate due process balance." 424 U.S. at 347.

21. See notes 31-33 *infra* and accompanying text.

22. 397 U.S. 254 (1970). The case involved a procedural due process challenge to the procedures in New York for terminating welfare benefits. The challenged programs included both Aid to Families with Dependent Children (AFDC), a federally assisted program of categorical assistance administered by the states, and New York state's general Home Relief Program, a program of general assistance funded and administered exclusively by New York state and local governments. *Id.* at 255-56.

23. *Id.* at 262-63 (citation omitted). The phrase "condemned to suffer grievous loss" was quoted from Justice Frankfurter's concurring opinion in Joint Anti-Fascist Refugee Comm. v. McGrath, 341 U.S. 123, 168 (1951).

24. Goss v. Lopez, 419 U.S. 565, 577-80 (1975) (suspension of high school students); Wolff v. McDonnell, 418 U.S. 539, 556, 560-61 (1974) (prison discipline); Arnett v. Kennedy, 416 U.S. 134, 167-68 (1974) (Powell, J., concurring) (firing of civil service employee); Morrissey v. Brewer, 408 U.S. 471, 481-84 (1972) (revocation of parole); Bell v. Burson, 402 U.S. 535, 540 (1971) (suspension of driver's license); Boddie v. Connecticut, 401 U.S. 371, 377-79 (1971) (inability to pay court costs for divorce proceeding). In Fusari v. Steinberg, 419 U.S. 379, 389 (1975) (unemployment compensation benefits), the Court did not decide the case on the merits, but it did reference the need for balancing the competing government and individual interests. In Board of Regents v. Roth, 408 U.S. 564, 569-70 (1972) (nonretention of university professor), the Court also discussed the need to balance these interests to determine the form of hearing required by due process. The Court did not actually balance the interests in *Roth* since it determined that due process protections did not apply in the respondent's case.

Goldberg.[25] Since then, with three exceptions,[26] the guidance from the Court has been minimal.[27]

A fundamental reason for the sparse development of the law in this area is that the balancing test heretofore enunciated by the Court is largely unworkable.[28] For the most part, the Court has repeatedly identified the same interests in all cases involving the termination of money benefits from the government. The articulated government interest is the conservation of administrative and fiscal resources.[29] The individual's interest is identified as the uninterrupted receipt of payments.[30] This process of identifying these competing interests and attempting to balance them does not lead to the conclusion that certain procedural safeguards must inevitably follow in specific cases whereas others do not.

The *Eldridge* opinion wisely tilts the due process balancing mechanism. The change does not abandon all consideration and comparisons of governmental and private interest. Rather, it simply raises the individual's interest to a level of paramount importance, shifting the emphasis of deciding what process is due to a primary focus on the individual:

> [M]ore is implicated in cases of this type than ad hoc weighing of fiscal and administrative burdens against the interests of a particular

---

25. 397 U.S. at 267-71.

26. Goss v. Lopez, 419 U.S. 565 (1975); Wolff v. McDonnell, 418 U.S. 539 (1974); Morrissey v. Brewer, 408 U.S. 471 (1972). Unlike *Goldberg*, none of these opinions involved government money payments.

27. The *Eldridge* majority opinion itself notes that "[i]n other cases requiring some type of pretermination hearing as a matter of constitutional right the Court has spoken sparingly about the requisite procedures." 424 U.S. at 333.

28. *See* Note, *Specifying the Procedures Required by Due Process: Toward Limits on the Use of Interest Balancing*, 88 HARV. L. REV. 1510, 1519-20 (1975). *See also* J. MASHAW & R. MERRILL, THE AMERICAN PUBLIC LAW SYSTEM 371 (1975) ("But even if properly understood the weighing process is excruciatingly difficult and inexact. For it is seldom clear precisely what costs and benefits are involved or what contribution various adjudicatory procedures will make to maximizing the net of benefits over costs."); Note, *Procedural Due Process and the Termination of Social Security Disability Benefits*, 46 S. CAL. L. REV. 1263, 1291-92 (1973) ("A pure balancing approach is inadequate to determine what procedural safeguards should be provided; rather, the courts must look to the truth ascertainment value underlying procedural due process to resolve the problem of 'how much process is due.'"). For a mixed assessment on interest balancing, see K. DAVIS, *supra* note 2, § 7.00-8, at 265.

29. Mathews v. Eldridge, 424 U.S. 319, 347 (1976); Arnett v. Kennedy, 416 U.S. 134, 168 (1974) (Powell, J., concurring) (reference to government's interest in terms of "the maintenance of employee efficiency and discipline"); Goldberg v. Kelly, 397 U.S. 254, 265-66 (1970).

30. Mathews v. Eldridge, 424 U.S. 319, 340 (1976); Arnett v. Kennedy, 416 U.S. 134, 168 (1974) (Powell, J., concurring); Goldberg v. Kelly, 397 U.S. 254, 266 (1970). The Court has pointed to this interest because of the provisions in the enabling statutes which both afford a claimant an opportunity for a hearing prior to a final administrative decision and entitle him to retroactive payments if ultimately successful.

> category of claimants. The ultimate balance involves a determination as to when, under our constitutional system, judicial-type procedures must be imposed upon administrative action to assure fairness.[31]

The Court, in *Eldridge*, assumes that the incremental cost burden of providing more hearings and of continuing benefit payments to ineligible recipients pending decisions would be substantial, but such costs alone are deemed "not a controlling weight."[32] The conservation of resources is still important in the Court's view because "[a]t some point the benefit of an additional safeguard to the individual affected by the administrative action and to society in terms of increased assurance that the action is just, may be outweighed by the cost."[33] Although the Court does not identify that transition point, its emphasis upon assuring fairness to the individual—particularly in light of its assumption that the fiscal and administrative burdens would be substantial—suggests that the governmental interest must have compelling weight in order to be a significant factor in determining the degree of due process to be afforded the individual.

By making protection of the individual the focal point of analysis, due process is approached in a manner more consonant with the tenets of the fifth and fourteenth amendments. The government can constitutionally make individualized decisions to affect or terminate life, liberty, or property, since these are not inalienable rights of a person.[34] But any governmental action which deprives a person of such interests must conform to due process of law. A straight balancing of the government's interest against that of the individual raises the governmental interest to a parity not intended by the Constitution. Due process represents the minimum level of protection by which the exercise of state power against the individual is to be controlled. In its procedural context, the question of what process is due is essentially one of ascertaining what minimum procedures will assure fairness to the individual so that his notice and opportunity to be heard are meaningful. The shift in emphasis evidenced in the *Eldridge* opinion recognizes this function and provides a better basis by which to achieve it.

---

31. 424 U.S. at 348. Notably the lower courts and other commentators have not made reference to this part of the *Eldridge* opinion. As is explained throughout the article, however, the direction indicated therein by Justice Powell is the direction which the Court should follow.

32. *Id.*

33. *Id.*

34. Such rights may be taken away, as when the state executes or incarcerates a criminal or fines an offender.

II. INADEQUACIES OF THE *Eldridge* ANALYSIS

Although the Court commendably reorients the general framework of analysis of procedural due process questions, its application of that analysis to the specifics of the *Eldridge* case and the implications which that analysis holds for future decisions raise specific problems. Each of these problems is separately considered in the subsequent parts of this section. These problems are fostered by the way in which the Court responds to both the articulated interest of the individual disability insurance recipient and the nature of the issues such termination cases raise.[35] The Court's conclusory position is predicated upon speculation and generalizations which patently disregard contrary facts. Unfortunately, its position is so encompassing that no money claims other than continued eligibility for welfare payments can realize the benefit of the Court's new emphasis of focusing on the individual's interest.[36] In short, the *Eldridge* analysis offers a role of increased importance for the private interest in the balancing equation, but at the same time dimin-

35. Most of the articles and notes on the *Eldridge* opinion have criticized it. *See* Mashaw, *The Supreme Court's Due Process Calculus for Administrative Adjudication in Mathews v. Eldridge: Three Factors in Search of a Theory of Value*, 44 U. CHI. L. REV. 28 (1976); Note, *Mathews v. Eldridge: Procedural Due Process Requirements in Social Security Disability Benefit Terminations*, 30 SW. L.J. 647 (1976); 60 MARQ. L. REV. 129 (1976); 8 ST. MARY'S L.J. 373 (1976). *But see* Note, *Social Security: Evidentiary Hearings Not Required Prior to Termination of Disability Benefits*, 6 CAP. U. L. REV. 327 (1976); 45 U. CIN. L. REV. 672 (1976).

36. Throughout this article the term "welfare" is used in a broad sense to include all welfare programs, not just those involved in *Goldberg*. See note 22 *supra*. A major eligibility requirement for welfare is financial need. Federal welfare programs include Supplemental Security Income, 42 U.S.C. §§ 1381-1395 (Supp. V 1975); Medicaid, 42 U.S.C. § 1396 (1970 & Supp. V 1975); Food Stamps, 7 U.S.C. §§ 2011-2025 (1970 & Supp. I-V 1971-75); and Aid to Dependent Children, 42 U.S.C. §§ 601-610 (1970 & Supp. V 1975). Additional state programs include general relief, emergency assistance, and medical benefits only.

Two subsequent circuit court opinions relied upon *Eldridge* to hold that evidentiary hearings are required prior to the suspension of supplemental security income payments since these benefits are based on financial need. Johnson v. Mathews, 539 F.2d 1111 (8th Cir. 1976); Tatum v. Mathews, 541 F.2d 161 (6th Cir. 1976). The District of Columbia Circuit also stressed that the Food Stamp Act is based on need in remanding a case in which the district court had dismissed a claim based on a due process challenge to termination of food stamp benefits. Basel v. Knebel, 551 F.2d 395 (D.C. Cir. 1977). In comparison, cases involving the termination of government money benefits other than welfare benefits have denied any due process requirement of a prior evidentiary hearing. Mattern v. Mathews, 427 F. Supp. 1318 (E.D. Pa. 1977) (recoupment for overpayments of disabled widow's benefits); Graves v. Meystrik, 425 F. Supp. 40 (E.D. Mo. 1977) (termination of unemployment compensation benefits); Stolberg v. Caldwell, 423 F. Supp. 1295 (D. Conn. 1976) (suspension of salary payments for services as college professor under "dual-job" ban). The court in *Mattern* succinctly articulates the distinction drawn by the *Eldridge* Court: "Now *Eldridge* holds that *Goldberg* is limited to welfare assistance given to persons on the 'very margin of subsistence,' involving the 'very means by which to live' and draws an even finer line of distinction between the needs of a social security recipient and a welfare recipient." 427 F. Supp. at 1328.

ishes the effect of that role with respect to the particular procedural protections which will be required. Not only is this position at odds with known facts in the area of disability payments, it will also predictably leave legislatures and administrative agencies disinclined to devise and implement needed procedures by which to direct these and other government terminations of money benefits.

A. Disregarding the Facts: Consequences of Error

The Court's approach to the interests of the disability insurance recipient in *Eldridge* is predicated upon a comparison with the interests of the welfare recipient in *Goldberg*. In both cases the government conceded the applicability of the due process clause to the property interest, created by statute, of persons qualified to receive the benefits.[37] Since the challenged procedures of both the welfare and the disability insurance programs provided for an administrative hearing *after* the termination of payments, and for retroactive payments in the event of an erroneous initial decision, the issue in both cases was narrowed to whether due process requires an evidentiary hearing *prior* to a suspension or temporary deprivation of benefits.

The affirmative answer of the *Goldberg* opinion was not repeated in the *Eldridge* case. Instead, the Court distinguished between welfare and disability insurance recipients, in part, on the basis of the supposed consequences to the individual of an erroneous initial decision. Presuspension evidentiary hearings were required for welfare recipients because of their "brutal need" or "grievous loss."[38] The eligible person on welfare relief, the Court reasoned, is dependent upon the monthly payments; erroneous decisions cutting off benefits to such recipients would leave them without support for their basic needs.[39] Although the majority in *Eldridge* recognized that the private interest of the individual in the uninterrupted receipt of disability insurance payments and the potential injury of an erroneous termination of those benefits are "similar in nature to

37. This position is in accord with the definition of property adopted by the Supreme Court for purposes of the due process clauses: "To have a property interest in a benefit, a person . . . must have . . . a legitimate claim of entitlement to it." Board of Regents v. Roth, 408 U.S. 564, 577 (1972). The Court stressed that property interests protected by the due process clauses are not found within the Constitution; "[r]ather, they are created and their dimensions are defined by existing rules or understandings that stem from an independent source such as state law—rules or understandings that secure certain benefits and that support claims of entitlement to those benefits." *Id.*

38. 424 U.S. at 340-43.

39. *Id.*

that of the welfare recipient in *Goldberg*,"[40] the Court nevertheless distinguished the two types of benefits. Welfare entitlement is based solely upon financial need. Since, however, entitlement to disability insurance is "wholly unrelated to the worker's income or support from many other sources,"[41] the Court considered "potential sources of temporary income"[42] to be a sufficient basis of differentation. Because of this difference, the Court concluded that the termination of disability benefits does not require the same presuspension procedures as does the termination of welfare benefits.[43]

The Court's reasoning is subject to challenge in several respects. First, although the Court suggests numerous private sources from which disability insurance beneficiaries may derive income,[44] these indentified sources are only potential and many beneficiaries have no such income.[45] In addition, the Court directs any worker whose income falls to poverty levels to apply for welfare, but this response will meet subsistance needs only if applications are quickly processed and funds are promptly forthcoming. Obtaining public welfare immediately is not always as easy as suggested by the Court.[46] Finally, and most importantly, with public money sufficient to meet only basic needs, an erroneously terminated disability recipient forced onto welfare can lose significant equity in certain possessions. Indeed, the dissenting opinion in *Eldridge* indicates that "because disability benefits were terminated there was a foreclosure upon the Eldridge home and the family's furniture was repossessed, forcing Eldridge, his wife and children to sleep in one bed."[47] In short, during the lengthy period between the time the

---

40. *Id.* at 340.

41. *Id.* at 340-41.

42. *Id.* at 343.

43. Professor Davis has concluded that this basis of distinguishing the cases is satisfactory. K. DAVIS, *supra* note 2, § 7.00-1-1, at 85 (Cum. Supp. 1977).

44. These are "earnings of other family members, workmen's compensation awards, tort claim awards, savings, private insurance, public or private pensions, veterans' benefits, food stamps, public assistance, or the 'many other important programs, both public and private, which contain provisions for disability payments affecting a substantial portion of the work force.'" 424 U.S. at 341, *citing* Richardson v. Belcher, 404 U.S. 78, 85-87 (1971) (Douglas, J., dissenting).

45. The Court noted that in 1965 the median income for disabled workers was only $2,836 and that median liquid assets totaled $940. 424 U.S. at 342 n.26. *See also* Brehm, *The Disabled on Public Assistance*, 33 SOC. SEC. BULL. 26, 29 (1970). Nearly sixty percent of the severely disabled men did not have any earnings in 1965; half of those employed earned less than $700. I. SWISHER, SOURCES AND SIZE OF INCOME OF DISABLED (social security survey of the disabled: 1966, Report No. 16, June, 1971).

46. *See* Meyerhoff & Mishkin, *supra* note 14, at 564-65.

47. 424 U.S. at 350. One of the plaintiffs in Richardson v. Wright, 405 U.S. 208 (1972)

disability insurance recipient is wrongfully cut off from payments to the time a post-suspension hearing is conducted in his case and a decision is finally made,[48] that individual can be very adversely affected.

Because of its willingness to disregard such facts, the Court has created a harsh constitutional standard.[49] It has limited the "grievous loss" characterization to situations in which a worker and his family are forced below the subsistence level. Rather than relating the due process requirement of a presuspension evidentiary hearing to a demonstrable actual serious loss, the *Goldberg/Eldridge* approach presupposes that such losses always exist in welfare cases, but precludes a similar showing of loss in any other case. If *Eldridge* is followed consistently in the future, a hearing prior to the suspension of money benefits will only be required in cases of welfare payments.[50]

(vacated as moot and remanded), had to sell his home because of loss of disability payments, although his benefits were later reinstated. *See* Meyerhoff & Mishkin, *supra* note 14, at 555, *citing* Brief for Appellees at 5-6, 28, Richardson v. Wright, 405 U.S. 208 (1972).

48. The Secretary of Health, Education, and Welfare conceded that:

> the delay between a request for a hearing before an administrative law judge and a decision on the claim is currently between 10 and 11 months. Since a terminated recipient must first obtain a reconsideration decision as a prerequisite to invoking his right to an evidentiary hearing, the delay between the actual cutoff of benefits and final decision after a hearing exceeds one year.

424 U.S. at 341-42.

49. The Court's language in *Goldberg* stressed the unique situation of the welfare recipient, suggesting that the requirement of a presuspension hearing would be limited to cases where welfare benefits are terminated:

> For qualified recipients, welfare provides the means to obtain essential food, clothing, housing, and medical care. . . . Thus the crucial factor in this context—a factor not present in the case of the blacklisted government contractor, the discharged government employee, the taxpayer denied a tax exemption, or virtually anyone else whose governmental entitlements are ended—is that termination of aid pending resolution of a controversy over eligibility may deprive an *eligible* recipient of the very means by which to live while he waits. Since he lacks independent resources, his situation becomes immediately desperate.

397 U.S. at 264 (emphasis in original). Two years later, however, in *Fuentes v. Shevin*, the Supreme Court indicated that such a narrow applicability of constitutionally required presuspension hearings was not intended: "While . . . *Goldberg* emphasized the special importance of . . . welfare benefits, [it] did not convert that emphasis into a new or more limited constitutional doctrine." 407 U.S. 67, 89 (1972). The majority in *Eldridge*, however, leaves no doubt that it intends to afford the protection of a constitutional standard only to the benefits peculiar to *Goldberg*.

50. The observation that a presuspension hearing will not be required does not apply to all cases involving a deprivation of government benefits, but is limited to those involving *money* benefits. The Court, in Goss v. Lopez, 419 U.S. 565 (1975), required notice and, in case of dispute, an abbreviated form of a hearing prior to "a suspension [of high school students] of ten days or less." *Id.* at 581. In a footnote to that holding, the Court rejected the judicial review available under the Ohio Code as an inadequate cure because the student would "irreparably lose his educational benefits" during the attendant delay. *Id.* at 581-82

The position of the *Eldridge* majority on the consequences of erroneous initial decisions in terminating non-welfare government money benefits is unfortunate. It does have the appealing advantage of ease of application with at least a modicum of rationale. If the due process right to a pretermination hearing depends upon use of a "grievous loss" standard, however, the Court should not protect just one category of persons because they are easy to ascertain and ignore all the others. A constitutional doctrine need not be so disjoined from reality. The real problem is to define the standard within realistic parameters and then to identify those persons who actually are encompassed within it. Before dealing further with that basic problem,[51] additional inadequacies of the *Eldridge* analysis must be considered.

B. Disregarding the Facts: Likelihood of Error

In analysing the nature of the issues presented in social security disability termination cases, the *Eldridge* Court again made a direct comparison with the *Goldberg* decision. *Goldberg* required the availability of cross-examination because the issues of witness credibility and veracity are often central to welfare decisions.[52] Because of these issues, and in addition, since most welfare recipients lack the education needed to express themselves well in writing, the *Goldberg* Court also mandated an opportunity for oral presentation.[53] In *Eldridge*, however, the Court characterized the determination of continued eligibility for social security disability insurance as a medical decision that turns principally "upon 'routine, standard, and unbiased medical reports by physician specialists.' "[54]

This approach oversimplifies the nature of many of the disability insurance termination cases. The problem with the Court's analysis stems from its excessive use of *Goldberg* as a foil. Once it distinguishes the social security cases on both of the criteria articulated in *Goldberg*, the requirements of cross-examination and oral presentation, the Court considers the contrast complete. In some disability cases, however, other issues are also raised which, contrary to the

---

n.10. The Court did not actually attach the label of "grievous loss," but it seems implicit in the decision. Unlike the situation involving a recipient of money benefits, a suspended student cannot be retroactively awarded the days of schooling he would lose if wrongfully suspended. Nor is an equivalence of welfare available during the suspension period since the student cannot turn to any other free educational institution to protect that interest.

51. See note 80 *infra* and accompanying text.

52. 397 U.S. at 269-70.

53. *Id.* at 267-69.

54. 424 U.S. at 344, *citing* Richardson v. Perales, 402 U.S. 389, 404 (1971).

Court's assertion, indicate that an evidentiary hearing, or at least an oral presentation, would be of substantial value in lessening the risk of an erroneous determination.

The distinctions the Court does raise are, for the most part, valid. The issue of credibility is involved in some of the disability cases,[55] but undoubtedly to a lesser extent than in welfare cases. Cross-examination is therefore less essential in disability determinations. Administrative procedures governing disability cases allow the recipient to examine the case against him and to submit evidence in rebuttal.[56] These procedures might, as the Court suggests, validly eliminate the *Goldberg* objections to written presentations in cases where the nature of the inquiry is similar to that in the welfare context or where only straightforward medical determinations are involved.

The Court fails, however, to adequately consider additional issues which are relevant in some social security disability cases.[57]

---

55. The SSA Claims Manual indicates that a continuing disability investigation initiates upon receipt of reports that a beneficiary has improved or returned to work. It states that "[o]ccasionally, information is received from persons wishing to remain anonymous." SOCIAL SECURITY ADMINISTRATION, CLAIMS MANUAL § 6705(a) (1975). The SSA is required to disclose such information to the beneficiary:

> [W]hen the evidence materially conflicts *in any way* with the beneficiary's assessment about his engaging in SGA [substantial gainful activity], the proposed date of cessation, entitlement to a trial work period, or where evidence is received from a 3rd party (e.g., employer), the beneficiary must be contacted by the DO [district office] to resolve the conflict (or to confirm 3rd party evidence) before "due process" is satisfied and a cessation determination can be prepared by the SA [state agency].

Id. at § 6705.2(b)(1) (1972) (emphasis added).

56. The Court viewed the administrative procedures as sufficient to assure that written submission will provide an effective means for the recipient to communicate his case to the decisionmaker. A detailed questionnaire is sent to recipients which "identifies with particularity the information relevant to the entitlement decision." The recipient can obtain personal assistance from the local Social Security Administration office in completing the questionnaire. 424 U.S. at 345. *See also* SOCIAL SECURITY ADMINISTRATION, CLAIMS MANUAL §§ 6705.1(b), .2(b)(1), .3(b) (1972). The recipient is also provided full access to information relied upon by the agency, as well as tentative assessments and explanations of reasons, prior to any suspension of benefits. The recipient is continually allowed the opportunity "to submit additional evidence or arguments, enabling him to challenge directly the accuracy of the information in his file as well as the correctness of the agency's tentative conclusions." 424 U.S. at 346. Medical evidence is provided by laboratory tests, X-rays and written reports from doctors.

The *Eldridge* Court placed particular emphasis on the significance of written medical evidence in making the disability determination. The Court cited with favor an earlier disability insurance benefits case, Richardson v. Perales, 402 U.S. 389 (1971). In *Perales*, "the Court recognized the 'reliability and probative worth of written medical reports,' emphasizing that while there may be 'professional disagreement with the medical conclusions' the 'specter of questionable credibility and veracity is not present.'" 424 U.S. at 344.

57. The subjective nature of some disability determinations is indicated in Underwood v. Ribicoff, 298 F.2d 850, 851 (4th Cir. 1962):

> [T]here are four elements of proof to be considered in making a finding of Claimant's

The Social Security Act provides that, in addition to medical diagnosis of impairment, the disability determination can also involve consideration of vocational factors, such as age, education, and work experience.[58] The Court makes an oblique reference to these issues, but states in conclusory form that "information concerning each of these worker characteristics is amenable to effective written presentation."[59] The *Eldridge* Court implies that all disability cases can be determined on objective criteria which are easily reducible to writing. But since the statutory criteria for disability must all be interrelated to a medical assessment of impairment, a case often involves a number of subjective evaluations.[60] The medical impairments must be projected in terms of functional limitations and correlated to such vocational factors as age, education, and work experience.[61]

The nature of these issues is such that the face-to-face contact of a hearing would be valuable in close cases. Also, the psychological condition of the recipient—a relevant decisional factor not conducive to determination exclusively by a medical expert—can be demonstrated at a hearing.[62] Several other kinds of cases can be identi-

---

ability or inability to engage in any substantial gainful activity. These are: (1) the objective medical facts, which are the clinical findings of treating or examining physicians, divorced from their expert judgments or opinion as to the significance of these clinical findings, (2) the diagnoses, and expert medical opinions of the treating and examining physicians on subsidiary questions of fact, (3) the subjective evidence of pain and disability testified to by Claimant, and corroborated by his wife and his neighbors, (4) Claimant's educational background, work history, and present age.

58. As the *Eldridge* Court noted, "the decision is not purely a question of the accuracy of a medical diagnosis since the ultimate issue which the state agency must resolve is whether in light of the particular worker's 'age, education, and work experience' he cannot 'engage in any . . . substantial gainful work which exists in the national economy . . . .' 42 U.S.C. § 423(d)(2)(A)." 424 U.S. at 344 n.28.

59. *Id.*

60. *See* Popkin, *Effectiveness of the Social Security Review System in Disability Cases*, 26 AD. L. REV. 79, 79-80 (1974).

61. *See* Mashaw, *supra* note 35, at 41-42, *citing* Haber, *Identifying the Disabled: Concepts and Methods in the Measurement of Disability*, 33 SOC. SEC. BULL. 17, 18-20 (1970).

62. The claimant's attitude towards enduring pain, discomfort and restricted movement, his enthusiasm for trying to learn a new skill, his willingness to adjust his lifestyle to the need for greater off-the-job rest, his residual vitality in general, and especially his pride, comprise his psychological 'set' and are relevant decisional factors in the borderline . . . area.

Dixon, *The Welfare State and Mass Justice: A Warning From the Social Security Disability Program*, 1972 DUKE L.J. 681, 708.

The Second Circuit, in Ber v. Celebrezze, 332 F.2d 293, 296 (2d Cir. 1964), considered the psychological condition of the recipient and found that

[w]hile the medical evidence may perhaps indicate that Mrs. Ber's physical symptoms were of a type which probably would have caused many people considerably less pain than Mrs. Ber suffered, it nevertheless amply supports her complaint that in her

fied where the nature of evidence is subjective.[63] As the required decision becomes increasingly subjective, the need for a hearing process by which to test it becomes more acute.

The opportunity for a presuspension hearing is, by no means, necessary in all social security disability cases. Many cases simply do not involve considerations of vocational factors[64] and not all of the remaining cases involve borderline decisions dependent upon subjective evidence.[65] As a threshold requirement, a worker must be able to demonstrate by means of "medically acceptable clincal and laboratory diagnostic techniques"[66] that he has a physical or mental impairment which is the cause of his continued "inability to engage in any substantial gainful activity."[67] A number of claimed disabilities are therefore resolved exclusively on the basis of this medical evidence.

The flaw in the Court's decision in *Eldridge* is that it treats all disability cases as if they fit a single mold. Just as it did in consider-

---

particular medical case these symptoms were accompanied by pain so very real to her and so intense as to disable her.

63. Cases involving "muscular or skeletal disorders, neurological problems, and multiple impairments, including psychological overlays" are examples. Mashaw, *supra* note 35, at 53. *See also* Liebman, *The Definition of Disability in Social Security and Supplemental Security Income: Drawing the Bounds of Social Welfare Estates*, 89 HARV. L. REV. 833 (1976).

> This focus on the reasonable expectations of Social Security participants suggests the need for fact-finding procedures open to allegations of pain; to assertions that a physical accident has led to changes in personality so drastic as to prevent effective work; and to claims that an individual is disabled even if he declines a dangerous and frequently unsuccessful operation.

Id. at 847.

64. Approximately one-half of the cases in the initial determination state are decided on the basis of objective medical evidence. "This wholly medical approach presumes disability if certain medically verifiable impairments are present. Cases in this category might be termed 'hard-core disability' and seldom present decisional problems which require appeal." Dixon, *supra* note 62, at 692 n.49. See note 65 *infra*.

65. Three tests of disability are included in the SSA regulation, 20 C.F.R. § 404.1502 (1971). These tests "form a progression from 'hard-core' to 'borderline' disability." Dixon, *supra* note 62, at 703. Professor Dixon has characterized them as an "objective medical approach" (Test I), § 404.1502(a), a "semi-objective standard" (Test II), § 404.1502(c), and the "subjective medical-vocational approach" (Test III), § 404.1502(b). *Id.* at 703-06. Test I leads to a finding of disability without consideration of vocational factors when the claimant's medical impairment correlates with either a category in a schedule of impairments or its medical equivalent. Test II allows a disability finding, despite some residual capacity to perform some gainful activity, where the individual is a poorly educated, older physical laborer without special skills, and he is unable to engage in lighter work. "Like Test I this criterion conflicts with the literal language of the federal statute, but apparently has congressional acquiescence." *Id.* at 705. Test III, stated essentially in the terms of the statute, provides for disability if impairments, in light of the vocational factors, make the claimant unable to engage in any substantial gainful activity.

66. 42 U.S.C. § 423(d)(3) (1970).

67. *Id.* § 423(d)(1)(A) (1970).

ing the *consequences* of an erroneous determination, the Court overgeneralizes about the nature of the inquiry when it considers the *likelihood* of erroneous determinations. This all-inclusive form of analysis disregards facts to the contrary which indicate that in some cases the effective communication of evidence of disability requires a personal appearance.[68] The root of the inadequacies of the Court's analysis in *Eldridge* is its willingness to overgeneralize.

C. *Creating Disincentives*

Legislation, whether it prohibits certain conduct or creates statutory entitlements, is not self-executing. Consequently, legislative bodies should concern themselves with more than just the substantive aspects of their enactments and, in particular, should consider the procedural requirements of new and existing programs. Administrative agencies, as the recipients of delegated powers from legislatures, should be equally concerned with the procedural ramifications of the programs under their control. Both entities possess the power to promulgate procedural, as well as substantive, rules; both should understand the need to accommodate competing interests in arriving at final determinations; and both have the flexibility to adapt procedures to fit particular circumstances.

The Court's application of the procedural due process standard simply measures existing procedures against the minimal requirements imposed by the Constitution. Through judicial review under that standard, however, the courts have the ability to press the legislature and the agencies to promulgate procedural rules that will adequately protect the individual. The Court's opinion in *Eldridge* is unlikely to create any such impetus for legislatures and administrative agencies. To the contrary, it is likely to act as a disincentive. Despite readily available facts indicating that many disability insurance recipients are forced into serious loss situations by the termination of their benefits pending agency hearings, the Court has concluded that only welfare recipients fit the "grievous loss" category,[69] and has assured that disability determinations will essentially be made only upon consideration of medical evidence.[70] The Court's approach relieves legislative bodies and administrative agencies from responsibility for devising methods to identify which

68. "Most of the cases which reach at least the hearing examiner level involve this broad test [Test III, see note 65 *supra*], because even a person with relatively severe impairments may have sufficient residual skill and physical capacity to perform a number of relatively simple occupations." Dixon, *supra* note 62, at 705.

69. See notes 33-47 *supra* and accompanying text.

70. See notes 48-66 *supra* and accompanying text.

recipients of government money benefits would, in fact, face serious loss if denied a presuspension hearing, and which cases involve issues in which a personal appearance would be helpful.

This disincentive extends beyond the lack of a positive requirement to undertake the task of improving procedures. It will also give pause to anyone motivated to improve agency procedures for cases of termination of disability benefits. Any procedures voluntarily adopted are likely to come under close scrutiny in judicial review. Legislatures and agencies can avoid the whole issue by simply continuing to use the existing procedures approved by the *Eldridge* Court.

D. Summary

The Supreme Court, in *Eldridge*, improved the position of the individual recipient in theory, but not in practice. It recognized that the interest of the individual is to be given paramount importance in the due process standard, but it confined that interest by overbroad, restrictive characterizations. The approach was demonstrably based upon inaccurate assumptions and excessive generalizations. To apply the statement of Professor Newman, the margin of error, as reflected in the likelihood and consequences of erroneous determinations, is simply too great with this approach. Unfortunately, the Court's position in *Eldridge* also creates a disincentive to narrow that margin of error.

III. Suggestions for Applying the Due Process Standard

A number of commentators have articulated various criteria for evaluating procedures.[71] One of the best discussions identifies the competing considerations of accuracy, efficiency, and acceptability.[72] The first two of these criteria correlate very closely to the major private and public interests which the courts have identified in applying the balancing test to decide what process is due in particu-

71. Boyer, *Alternatives to Administrative Trial-Type Hearings for Resolving Complex Scientific, Economic, and Social Issues*, 71 Mich. L. Rev. 111, 137-50 (1972); Cramton, *A Comment on Trial-Type Hearings in Nuclear Power Plant Siting*, 58 Va. L. Rev. 585, 591-93 (1972); Davis, *The Requirement of a Trial-Type Hearing*, 60 Harv. L. Rev. 193 (1956); Robinson, *The Making of Administrative Policy: Another Look at Rulemaking and Adjudication and Administrative Procedure Reform*, 118 U. Pa. L. Rev. 485, 521-22, 536 (1970); Rosenberg, *Devising Procedures that are Civil to Promote Justice that is Civilized*, 69 Mich. L. Rev. 797, 802-03 (1971); Rubenstein, *Procedural Due Process and the Limits of the Adversary System*, 11 Harv. C.R.-C.L.L. Rev. 48, 65-76 (1976); Verkuil, *A Study of Informal Adjudication Procedures*, 43 U. Chi. L. Rev. 739, 742-57 (1976).

72. Cramton, *supra* note 71; *accord*, Boyer, *supra* note 71.

lar cases. The case law has most often related the private interest to consideration of accuracy so as to avoid the likelihood and consequences of erroneous determinations. The Court articulated governmental interest of conserving administrative and fiscal resources closely corresponds to the criterion of efficiency.[73] This correlation of private and public interests to considerations of accuracy and efficiency, in light of the *Eldridge* emphasis on the individual's interest, suggests that accuracy has the role of paramount importance over efficiency from the perspective of constitutional requirements. Ideally the procedures utilized should advance both types of interests and serve the goal of accurate and efficient determinations. The difficult task in melding these two criteria, as indicated at the beginning of this article, is to ascertain the tolerable margin of error.

We should not look, however, to the concept of procedural due process to draw that line. Due process decisions are not the proper implement by which to make close refinements between the competing considerations of accuracy, efficiency, and acceptability. Due process is a position of last resort; it is the ultimate guarantee of minimum procedural protection by which governmental action directed toward individuals must be measured. The due process clauses are written in terms of protection of the individual, and the Supreme Court in *Eldridge* properly identifies the individual's interest as the paramount concern. If other methods are not utilized to approach the permissible margin of error with greater precision, due process should dictate procedures which will assure safeguards for the individual, even if such a guarantee leads to some "excess" procedure.

To the extent that excess procedures are undesirable, as they are, at least from an efficiency perspective, the excess should be pared away by governmental bodies other than the courts. The responsibility lies with legislative bodies and administrative agencies. Rulemaking and administrative bodies should direct their initiative and innovation toward developing procedures that will satisfy the various criteria by which procedural systems are measured and adjudged sound. When their failure to meet that challenge leads to infringements upon the basic procedural protections to be afforded an individual, the courts must then compel adherence to that basic standard.

Since the legislatures and agencies cannot or simply will not do the task alone, they need both prodding and assistance from the

73. The government also has some interest in accurate decision-making and individual parties are concerned about efficient procedures. *See* Verkuil, *supra* note 71, at 740, 743.

judiciary.[74] Both of these levels of judicial participation are discussed in the remainder of this article. When the Court approves existing procedures in the face of apparent deficiencies, as it did in *Eldridge*, it does not provide the necessary incentives. On the other hand, when the Court, in other cases, has specified particular procedures that will be required, the result has also been to remove the incentive for further development of procedures.

A. Needed Judicial Prodding

The *Eldridge* Court would have acted more consistently with its principle of minimum protection for the individual if it had overturned the agency procedures and, on an interim basis, required presuspension hearings for all recipients of social security disability insurance. This requirement admittedly would have created excess procedure. It would have extended major procedural protections to a large number of persons whose cases could be decided solely on medical reports or who would not suffer actual serious losses if erroneously terminated. Nevertheless, the Court should have followed this action, since the procedural protection of presuspension hearings is needed by a significant number of the affected class of individuals,[75] and since the responsible agency had not taken steps to develop procedures that would enable it to identify those individuals with some degree of precision.

Such an interim requirement with its excess procedures would have assured protection for the individuals who needed it, and could have prompted the agency to devise techniques to legitimately remove the excess. The SSA could classify its cases, afford adequate procedures for each class, and thereby avoid excessive hearings. Any case that can be determined on the basis of the medical evidence alone should not, by the nature of the case itself, trigger a require-

---

74. The importance of the Court's pressing other governmental bodies to the task was noted in section II. C. *supra*.

75. Statistics prepared for a staff report on the disability insurance program indicate that in 1975 over 26% of the disability allowance determinations involved consideration of vocational factors.

> Although the 1967 legislation re-emphasized the medical factor as of predominant importance, experience over recent years shows that more and more cases are being determined on the nonmedical factors which are those which are the most subjective.
>
> Also, a large percentage of the disallowances involve evaluation of nonmedical vocational factors so that in all, 45 percent of all substantive determinations involve such factors, and this type of case represents a high percentage of those cases on appeal.

STAFF OF SUBCOMM. ON SOCIAL SECURITY OF THE HOUSE COMM. ON WAYS AND MEANS, DISABILITY INSURANCE—LEGISLATIVE ISSUE PAPER 17 (May 17, 1976).

ment for a presuspension hearing. The SSA could review past case histories to ascertain which types of cases have been the most susceptible to erroneous initial decisions. Cases which involve considerable subjective judgment or borderline determinations could thus be identified.[76]

An interim requirement for presuspension hearings could also encourage the SSA to develop procedures that will increase the accuracy of decisionmaking. The quest for accuracy depends initially upon the meaning of or objective envisioned by the concept. As the *Eldridge* Court correctly noted, "procedural due process rules are shaped by the risk of error inherent in the truthfinding process as applied to the generality of cases."[77] The goal in disability cases is not absolute truth or accuracy.[78] Rather the basic purpose is to assure the individual the opportunity for an effective means by which to present his case to the decison-maker. In borderline disa-

76. The Social Security Administration already has the three tests of disability. See note 65 *supra*. "In percentage terms the residuum of highly discretionary borderline claims under Test III is a small portion—15 to 20%—of the total intake but the residuum is large in absolute terms because of the high intake volume." Dixon, *supra* note 62, at 706 n.114.

77. 424 U.S. at 344.

78. Indeed, Professor Mashaw suggests that in the disability system "'accuracy' is meaningless and consistency unachievable or at least unachieved." Mashaw, *supra* note 35, at 45. In support of this observation he cites and discusses the high reversal rate on appeal of disability determinations and the lack of consistency in state agency disability determinations. *Id.* at 43-45. Mashaw's indictment should not, however, be considered sufficient to abandon the quest for accurate decisions. Reasons, other than inaccuracy, have been advanced which plausibly explain at least some of the causal factors of the high reversal rates. First, the Social Security Administration operates under an "open file-continuing claim concept [that] operates to inflate the reversal rate." Dixon, *supra* note 62, at 693-94 n.58. Second, a finding of disability on appeal may "simply reflect the worsening of a previously nondisabling impairment." Mashaw, *supra* note 35, at 43. Furthermore, only cases involving determinations of no disability receive further review. Dixon, *supra*, at 692 n.52. This skewed case population creates a psychological factor for the administrative law judge; "his impulse to allow a certain number of claims and his interest in exercising hearing examiner independence can be fulfilled only by reversing Test III [see note 65 *supra*] denials." *Id.* at 708. Finally, decisions made at the state agency level and the appeal level are based on different disability standards. *Id.* at 706-07; Mashaw, *supra* note 35, at 43-44. In order to improve the consistency of decision making it has been suggested that new statutory standards be promulgated. "The critical problem in the SSA disability program may center not in the several-tiered administrative determination process, but rather in the inexactness of the statutory standards, the lack of clarity and precision in the rule-making procedures, and the interrelationship of the standards problem with the administrative process." Dixon, *supra* note 62, at 700-01. *See also id.* at 687, 707, 710, 714, 718. Professor Davis has taken the position that "[t]he focus of judicial inquiries thus should shift from statutory standards to administrative safeguards and administrative standards." Davis, *A New Approach to Delegation*, 36 U. CHI. L. REV. 713, 713 (1969). He has also pointed out the willingness of some courts to correlate such a requirement to the concept of due process: "Some courts have already ignored the absence of statutory standards and have held that due process forbids the administrators to exercise their discretionary power in particular cases without first creating administrative standards or guides." *Id.* at 730.

bility cases where observation of the recipient is relevant, or where significant evidence is subjective, a personal appearance by the recipient may be essential.[79] An appearance will not generally make difficult issues vanish, nor will it eliminate the element of discretion which lies with the decision-maker. It does, however, recognize an important element of accurate determinations: the requirement that the exercise of discretion be based upon adequate information. In this sense, accuracy is a major objective in reducing the margin of error to a tolerable margin.

The SSA can develop procedures to identify individuals who would suffer a serious loss if the termination of disability benefits should ultimately prove to be erroneous. The agency could prepare a form that it would require all recipients to complete as a sworn statement prior to the suspension of benefits. This form should identify with particularity all potential sources of income and support, and inquire into the amounts that would be available to the recipient if his disability insurance payments were cut off pending the agency hearing. On the basis of these sworn statements, the agency could ascertain the financial position of the recipients and determine the necessity of a pretermination hearing.

The Court should also adopt a more realistic definition of "grievous loss" and should encourage its broad interpretation in order to make that standard viable. The Court's articulation of the private interest involved in cases of government payments—the uninterrupted receipt of those payments—is quite perceptive. Since eventual success on the merits statutorily entitles the recipient to retroactive payments, the loss to the erroneously suspended recipient relates not to the non-receipt of the money, but to the consequences of not receiving the money at the scheduled time. The problem with the Court's application of the "grievous loss" standard to that interest is that it is overly restrictive. Recipients other than those who are forced to a subsistence level will, in fact, be grievously affected by an erroneous determination.

Only certain easily identifiable recipients need to be added to the "grievous loss" category to improve its efficacy immeasurably. Those who will be forced to lose equity in non-liquid assets suffer a loss which will not be made up by subsequent retroactive payments. Although the total amount of money received from the government is the same, unless payments are made on time, assets in which

79. "None of these combined fact-law considerations can be readily determined by a medical expert, although medical factors are relevant, nor can they be precisely decided by a vocational expert." Dixon, *supra* note 62, at 708. See note 62 *supra* and accompanying text.

considerable equity may have accumulated can be repossessed.[80] In addition, the suspended recipient who relied exclusively upon disability benefits to meet his subsistence requirements will be unable to qualify for welfare benefits until his assets are reduced to the maximum exclusion level. Thus, a suspended recipient who owns a modest home but has no income or other assets would be forced to sell the home in order to have money for the basic needs of himself and his family.[81] The loss in such cases is major and exceeds what the individual should have to bear for an erroneous governmental determination. In a society so involved with credit transactions, recipients with a substantial interest in non-liquid assets should receive additional protection. Recipients forced to use liquid assets do not face such irreplaceable loss. Money from a savings account, whether it is used to provide for basic necessities of life or to continue monthly payments for a mortgage or installment transaction, can be replaced in lump sum upon receipt of retroactive payments. In the context of termination of government money benefits, "grievous loss" should therefore be redefined to include not only the consequence of being forced to a subsistence level, but also of being forced to lose equity in non-liquid assets.

This new definition of "grievous loss," and the classification of cases according to the risk of erroneous decision and the need for adequate information would greatly improve the procedures for the suspension of social security disability payments without mandating presuspension hearings in all cases. If a particular case can be determined solely on medical evidence or is one in which the nature of the issues is such that the risk of an erroneous determination is not high, a prior hearing is not necesssary due to the relatively high degree of reliability of the initial decisions. Cases with a greater likelihood of error suggest the necessity of a prior hearing. The only additional cases in which a presuspension hearing should be required are those in which the issues raised suggest a high risk of error *and* in which the consequences of such error meet the redefined criteria of the "grievous loss" standard. Such a procedural approach would be consistent with the broader due process protection for the individual that was recognized, but not applied, by the *Eldridge* Court.

---

80. *See* Morris, *Welfare Benefits as Property: Requiring a Prior Hearing*, 20 AD. L. REV. 487 (1968).

81. For example, in Ohio the maximum exclusion for a home in order to remain eligible for general relief is $12,000 market value after deducting liens. OHIO DEPARTMENT OF PUBLIC WELFARE, PUBLIC ASSISTANCE MANUAL § 325.2(1) (July 1, 1976). The total value of personal property may not exceed $2,250. *Id.* at § 326(2).

B. Needed Judicial Assistance

In addition to creating incentives for legislative bodies and agencies to adopt such techniques in their procedures, the Court should allow these entities sufficient leeway to act creatively upon such incentives. In prescribing various procedures which the Court believes constitute minimum needed safeguards, it has often dictated the procedures as absolutes, applicable across the board to all cases of the kind under review. These enumerated procedures are important because they provide immediately applicable protections for the affected individuals. They should be viewed, however, as an impetus to further development of procedures, and not—as they presently are—as the end of the inquiry.

By requiring particular procedures as absolutes, the Court impedes any further development of procedures and the chance for an improved accommodation of the values of accuracy, efficiency, and acceptability.[82] The Court should articulate some minimal procedures, but should also expressly indicate that other procedures which will provide the same level of protection will also satisfy the requisites of due process. Legislative bodies and agencies would then have the freedom to substitute other procedures for those suggested by the Court, as long as the basic guarantee of fairness to the individual was maintained.[83]

Such an approach to decision writing does involve a risk to eventual realization of the goal of procedural systems that represent a good accommodation of competing values. An agency may simply adopt the court-directed procedures without making any effort to seize the initiative to refine them further. If, however, the courts frankly recognize in their opinions that their articulated procedures may impose some degree of excessive protection and openly encourage the development of substitute procedures that approach the tolerable margin of error with greater precision, the motivation to take the initiative would be harder to resist. In addition, courts can force the hand of agencies somewhat by failing to set procedures in too much detail.[84]

---

82. Professor Mashaw emphasizes the acceptability aspect of evaluating agency procedures and suggests that "the Court develop a qualitative appraisal of the type of administrative decision involved." Mashaw, *supra* note 35, at 51. He also recognizes, however, that concern for acceptability values does not inevitably lead to the conclusion that presuspension hearings are required when a full hearing is later available: "Knowledge that an oral hearing will be available at some point should certainly lessen disaffection and alienation." *Id.* at 57. For additional discussion of acceptability, see Verkuil, *supra* note 71, at 752-57.

83. *See* Friendly, *supra* note 3, at 1278-79.

84. A suggestion by Professor Davis seems particularly relevant in this regard: "Perhaps the courts should refrain from formulating the precise procedures they find desirable; they

Finally, the Court should begin looking beyond a hearing or even judicialized procedures as the exclusive means to assure procedural due process. A growing number of commentators have proposed alternative means to assure fairness in the administration of government programs.[85] The *Eldridge* opinion does include some language that suggests in the future the Court may be willing to look beyond hearing procedures:

> We reiterate the wise admonishment of Mr. Justice Frankfurter that differences in the origin and function of administrative agencies "preclude wholesale transplantation of the rules of procedure, trial and review which have evolved from the history and experience of the courts." The judicial model of an evidentiary hearing is neither a required, nor even the most effective method of decision-making in all circumstances.[86]

Although this language admittedly may not embrace all alternative methods of decision-making, hopefully it will represent a prelude to formal recognition by the Court of additional means to satisfy the requirements of procedural due process.

IV. CONCLUSION

The *Mathews v. Eldridge* opinion is commendable for its position on the general framework of analysis of procedural due process, but unsatisfactory in its application of that analysis. The individual facing a potential government deprivation of a liberty or property interest appears to have finally secured basic minimum protection through the tilt of the due process balancing mechanism in favor of the individual. Unfortunately, the Court undermines the realization of this minimum protection. By disregarding facts to the contrary and overgeneralizing, the Court determines that neither the likelihood nor the consequences of error are high in social security disability cases. This decision creates disincentives for legislative and

---

should lay down broad minimum requirements, without writing codes of procedure." K. DAVIS, *supra* note 2, § 7.00-8, at 265-66; *accord*, Friendly, *supra* note 3, at 1301-02; Mashaw, *supra* note 35, at 29.

85. *See* Boyer, *supra* note 71; Friendly, *supra* note 3, at 1289-91, 1316; Frug, *Does the Constitution Prevent the Discharge of Civil Service Employees?*, 124 U. PA. L. REV. 942, 997 (1976); Kirp, *Proceduralism and Bureaucracy: Due Process in the School Setting*, 28 STAN. L. REV. 841, 870-71 (1976); Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772 (1974); Rabin, *Job Security and Due Process: Monitoring Administrative Discretion Through a Reasons Requirement*, 44 U. CHI. L. REV. 60 (1976); Verkuil, *supra* note 71, at 744-45.

86. 424 U.S. at 348, *quoting* FCC v. Pottsville Broadcasting Co., 309 U.S. 134, 143 (1940).

administrative bodies to take steps to correct apparent deficiencies in the procedures of this and other government programs. Due to the numerous shifts in the Court's due process analysis during this decade, consistency in its approach is clearly desirable. The problems with its application of due process in *Mathews v. Eldridge*, however, leaves too wide a margin of error to be acceptable.

Lying, Confidentiality, and the Adversary System of Justice

*Robert P. Lawry**

The legal profession has been indebted to Monroe Freedman since those stormy days in the mid-sixties when he was vilified and nearly disbarred for stating his arguments and conclusions about the demands that the "adversary system" places upon a lawyer's ethics.[1] His recent book on the subject[2] now assumes an indispensible place in the literature of professional responsibility, and it is, at present, the port for all departures. Dean Freedman's most significant contribution has been his emphatic reminder to keep the demands of the "system" constantly before us as we attempt to define the lawyer's ethical role within that system. He has thus challenged us to make a deeper and more careful examination of the adversary system than has ever been made before. After Freedman, any discussion of lawyers' ethics must proceed within the context of some stated definition or description of the adversary system.[3]

This admonition is more easily stated than followed because it is so difficult to probe one's "idea" of something he assumes he knows and understands—in this case, the adversary system of justice. In light of so much recent literature on the subject of lawyers'

* Associate Professor of Law, Case Western Reserve University School of Law.

1. *See* M. FREEDMAN, LAWYERS' ETHICS IN AN ADVERSARY SYSTEM viii (1975). The original controversy was sparked by a lecture, later published in Freedman, *Professional Responsibility of the Criminal Defense Lawyer: The Three Hardest Questions*, 64 MICH. L. REV. 1469 (1966).

2. M. FREEDMAN, *supra* note 1. The book as been extensively reviewed. *See, e.g.*, Dorsen, 11 HARV. C.R.-C.L.L. REV. 764 (1976); Douglass, 14 HOUSTON L. REV. 519 (1977); Kennedy, 4 FLA. ST. U.L. REV. 171 (1976); Kunstler, 4 HOFSTRA L. REV. 895 (1976); Lazarus, 51 N.Y.U. L. REV. 348 (1976); Margolis, 50 CONN. BAR J. 103 (1976); McDonald, 62 A.B.A. J. 848 (1976); Meagher, 4 FORDHAM URBAN L.J. 238 (1976); Neal, 29 VAND. L. REV. 529 (1976); Noonan, 29 STAN. L. REV. 363 (1977); Phillips, 43 TENN. L. REV. 179 (1975); Rotunda, 89 HARV. L. REV. 622 (1976); Singly, 38 U. PITT. L. REV. 135 (1976); Weckstein, 3 CIV. LIB. REV. 72 (No. 2, 1976); Weiss, 44 GEO. WASH. L. REV. 202 (1975).

3. In dealing with legal words and ideas, the quest for a classic definition *per genus et differentiam* has been criticized as futile. *See* J. BENTHAM, A FRAGMENT ON GOVERNMENT 106 n.1 (W. Harrison ed.) (1960). *See also Hart,* DEFINITION & THEORY IN JURISPRUDENCE (1953). Professor Hart has done much to continue and strengthen Bentham's attack, particularly in the work cited above and in his *The Concept of Law* (2d ed. 1966). However useful these criticisms may be, it still remains necessary to provide descriptive characteristics of things in order to talk about them meaningfully at all. The substitution of a paradigm case in place of a classic definition may not even amount to a noticeable change. My point is simply that one needs to understand what is meant by the use of the phrase "the adversary system;" whether one elucidates a concept by use of a paradigm or resorts to a classic definition is not relevant for purposes of this article.

ethics, the time seems propitious for each one concerned with the subject to lay bare his or her own "ideas" or "assumptions"[4] about the system. A modest start is necessary. Thus, this article will join in the Freedman debate, examining the demands the adversary system makes upon a lawyer's ethics, in light of a single question: What should a lawyer's response be when he or she "knows" that a client has lied, is lying, or will lie during the litigation process?[5] My proposed solution to this question will be diametrically opposed to that offered by Dean Freedman. My thesis is simply that no lawyer may allow a lie to corrupt the adversary system. Before advancing arguments in support of this thesis, however, it is incumbent upon me to lay bare my own general ideas and assumptions concerning the adversary system.

I. LYING AND THE ADVERSARY SYSTEM

A. *The Adversary System*

From a philosophical perspective, the adversary system has many analogues. Plato's dialogues and the scholastic *disputatio* are but two. It might be argued that Plato himself believed that truth emerges as a human reality *only* in dialogue, in conversation: "By conversing many times and by long, familiar intercourse for the matter's sake, a light is kindled in a flash, as by a flying spark."[6] The rules of the scholastic *disputatio* look much like procedural rules governing the trial of an issue: "To every *disputatio legitinia* there belongs question, answer, thesis, agreement, negation, argument, proof and concluding formulation of the result."[7] In the hands of a Thomas Aquinas, the adversary is treated with the highest respect, his arguments being formulated more cogently than he himself was able to do. It was the initial requirement of this kind of dialogue that a participant "first repeat the opposing objection . . . thus explicitly making sure that he fully understood what his oppo-

4. My concern is twofold: First, one's "image" of the system, for example, as a mere substitute for trial by battle, may work subsconsciously to distort one's view of what the system theoretically is or actually has become in practice. Second, this "image" will lead to the formation of propositions about the system that are less and less connected to the actual complex reality. No one can study philosophy for long and not emphasize the need for clearing the intellectual air of the old debris that hampers all fresh thinking on a subject. In this regard, see R. UNGER, KNOWLEDGE AND POLITICS 1-3 (1975) on the need for "total criticism."

5. Quite a lot depends upon the understanding of "timing" in the question being investigated. See note 122 and text accompanying note 120 *infra*.

6. PLATO, Seventh letter, 341 c, *quoted in* J. PIEPER, GUIDE TO THOMAS AQUINAS 74 (1962).

7. 2 M. GRABMANN, SCHOLASTISCHE METHODE 20, *quoting* Magister Radulfus.

nent had in mind."[8] With the issue joined, the arguments could be sharper. In all of this, there is an underlying conviction that "truth is an affair that calls for more power than the autarchic individual possesses."[9] Certainly truth will not always be achieved by a one-sided investigation into the "facts," and perhaps it is never achieved by one person's exposition. It takes the cutting edge of dialogue, the thrust and parry of two intellects in vital exchange, in order to crystallize a question and distill an answer. I believe the ideas embedded in these analogues, which I propose to call the "philosophy of dialogue," are also deeply embedded in the adversary system of justice. In his important writings on the adversary system as a philosophy of adjudication, Professor Lon Fuller apparently concurs.[10]

Fuller emphasizes that the system includes both the manner in which the trial should be conducted and the roles to be played by judges, juries, and advocates. Essentially, the advocates are to be partisan, to see for and speak for the party each represents. The trier of fact (judge or jury) is to remain above the fray, not to be involved in gathering facts, not to participate actively in the trial.[11] The idea is not borrowed from classical capitalist economic theory, however easily certain kinds of comparisons can be made.[12] Moreover, the adversary system is not merely a substitute for trial by battle, though again, there are others who seem to think so.[13] The idea is basically Platonic or Scholastic; it is rooted in the belief that the dialogue is at the heart of truthseeking.

In my judgment, as a system of adjudication, the adversary model is also based on at least two psychological assumptions[14]

---

8. J. PIEPER, *supra* note 6, at 77.

9. *Id.*

10. *See* Fuller & Randall, *Report of the Joint Conference*, 44 A.B.A.J. 1159 (1958) [hereinafter cited as *ABA Report*]; L. FULLER, THE ADVERSARY SYSTEM, TALKS ON AMERICAN LAW 30 (rev. ed. H. Berman 1971) [hereinafter cited as TALKS].

11. *See ABA Report, supra* note 10, at 1160-61; TALKS, *supra* note 10, at 30.

12. *See* Frankel, *From Private Fights Toward Public Justice*, 51 N.Y.U. L. REV. 516, 516-22 (1976). Judge Frankel calls the adversary system "The Adam Smith System of Adjudication."

This is *not* to say that general political, economic and social ideas do not influence the adversary system, and vice-versa. Clearly they do. How much, however, is a question no one appears capable of resolving. We just guess, argue and cite what evidence will support our arguments. See R. UNGER, *supra* note 4, for a fresh and bold attempt to guess and argue on a remarkably abstract level about these kinds of interconnections.

13. I would place Charles Curtis in this category. *See* Curtis, *The Ethics of Advocacy*, 4 STAN. L. REV. 3, 5 (1951) (a lawyer, on behalf of his client, "is required to treat outsiders as if they were barbarians and enemies").

14. I would say both these assumptions are consistent with the philosophy of dialogue sketched in the text, but neither are necessarily demanded by that philosophy.

about the way men conduct their affairs and the manner in which they reach fair and impartial judgments. These assumptions, in turn, are based on an intuitive realization that a sophisticated legal system is so complex, and man's abilities and resources so unevenly distributed, that trained representatives must be permitted to act for individuals who are attempting to deal with it. This intuition comes to all civilized societies, it seems, however adverse they are to lawyers (or their "advocate" analogues).[15] The two basic assumptions are: first, a partisan and highly trained advocate will work more diligently and will thus uncover more relevant facts than a neutral party. The advocate will also make the best possible arguments for his side of the dispute. Second, the judge will render a more fair, more impartial, more enlightened decision by remaining aloof from any independent investigation into the facts and by remaining largely passive while advocates for each side make their factual presentations and present their arguments.[16]

Although Professor Fuller does not identify these two assumptions, as such, in his writings, he does provide rationales for each.[17] Thus, behind the first assumption lies a conviction that partisan advocacy is the best way to sharpen issues so that a public trial of law and fact can be had[18]—a goal believed to be of independent value. Moreover, partisan advocates are less likely than a judge-administrator to become entangled in bureaucratic red tape that may both prolong the proceedings and blur the issues.[19] Behind the second assumption lies a goal that is also considered of independent value in society. As Fuller expresses it: "The institution of advocacy is not a concession to the frailties of human nature, but an expression of human insight in the design of a social framework within which man's capacity for impartial judgment can attain its fullest realization."[20]

Thus we can see that a "public trial of fact and law" and the desire to have "impartial judgment" are also goals of the adversary

15. R. POUND, JURISPRUDENCE 673-76 (1959).

16. Fuller argues that:

> An adversary presentation seems the only effective means for combatting this natural human tendency to judge too swiftly in terms of the familiar that which is not yet fully known. The arguments of counsel hold the case, as it were, in suspension between two opposing interpretations of it. While the proper classification of the case is thus kept unresolved, there is time to explore all of its peculiarities and nuances.

ABA Report, supra note 10, at 1160.

17. *Id.* at 1161.

18. *Id.* at 1160.

19. This point is my own. Fuller's concern with bureaucracy is that it produces the "too swift judging" he criticizes in the quotation at note 16 *supra*. TALKS, *supra* note 10, at 44.

20. *ABA Report, supra* note 10, at 1161.

system as we know it; it must never be forgotten, however, that underpinning the system is a basic desire to attain the "truth" and that the philosophy of the dialogue is therefore at the heart of the matter. As an unhappy by-product of the attempt to achieve other goals, it sometimes happens that a loss of truth occurs. Yet that loss is never deliberate. The fact that the system may result in a failure to gain the whole truth does not mean that the system can or does countenance lying. That point must be made again and again. When we discuss constitutional rights or the tradition of confidentiality between lawyer and client, this position must meet the challenge of adjusting to additional values.[21] I will argue that the adjustment may never include the acceptance of a lie by client or lawyer.

B. Two Fundamental Propositions

However the "adversary system of justice" may ultimately be defined or described, there are two fundamental propositions which are central to the system. They are: (1) no one has a right to lie under any circumstances within the system; and consequently, (2) no lawyer may countenance lying or cooperate with anyone in his or her attempt to lie. A close examination of these points will be helpful in resolving the question addressed in this article.[22]

1. *No Right to Lie*—For purposes of this discussion, the term "lie" shall be defined as "an untrue statement made with the intent of deceiving."[23] The philosopher may find this definition unacceptable for his or her own particular use,[24] but the definition comports with the common understanding of the word. This definition of the term is chosen in light of the adversary system. Where a "half-truth" may be a "lie" to a moralist, such a statement may not be a lie within the context of the adversary system, even though made with the intent to deceive or to alter a line of questioning. Let me explain.

The first thing to realize about this definition of lying is that a

---

21. See notes 56-72, 108-19 *infra* and accompanying text.

22. See note 5 *supra* and accompanying text.

23. This comports with the ordinary dictionary definition. FUNK & WAGNALL, STANDARD COLLEGE DICTIONARY 781 (1963); WEBSTER'S SEVENTH NEW COLLEGIATE DICTIONARY 487 (1965).

24. G.E.M. Anscombe, for example, defines a lie as "an utterance contrary to one's mind." G. ANSCOMBE, INTENTION 4 (1957). Thus a statement may be true (X is in Pittsburgh) though the speaker believes it to be false (believes X to be in Cleveland). Other interesting possibilities abound, particularly the more subjective one becomes. If one knowingly utters a true statement with the intent to deceive, there may also be good reason to call that statement a lie. One may *know* the hearer will interpret the statement differently from what its propositional form would otherwise indicate, or would even "clearly indicate" to 99 people out of 100.

"lie" is a positive statement. Silence will not be considered a lie; indeed, the right to remain silent in the criminal context under the fifth amendment has its civil analogue in the right to withhold information or to attempt to block its disclosure through a number of legally permissible routes. There is no lawyer who has practiced law for a modicum period of time who has not consciously used his or her lawyering skills to prevent access to information which the other side wants or would want if its existence were known.

Moreover, to be a lie, the statement itself must be "untrue," not one of the many variants of imaginable "half-truths." For example, suppose in a trial for price-fixing under the antitrust laws, a witness were asked whether he attended various meetings where "prices were discussed." If it were true that he never attended any meetings with other defendants where prices were discussed, he may answer "no" to any questions about attendance, even though he was in telephone contact with one or another person present at the meeting. Much depends on the form of the question in these matters and more depends on the skill and thoroughness of the lawyer's questioning. We may or may not believe in this approach, but we do behave that way, and that behavior seems woven into the fabric of the adversary system as we know it.

Judge Marvin Frankel would like to put an end to this "evasion" of truth by requiring each lawyer to elicit the "whole truth" from his own witnesses.[25] We have not yet taken Frankel's road, but not because we do not seek the unvarnished truth; we simply do not know how to achieve it, given other goals, like a "public trial," or "confidentiality," or "impartial judgment." So far, we have found it practically difficult to go Frankel's way. We depend on broadened discovery and on skillful cross-examination to attain the fullness of truth.[26] But mark how little even these remedies would avail, if they were not all premised on the basic proposition that no one may "lie."

In summary, it is clear then that "withholding information" is permissible under the "system," unless there has been an appropriate discovery request or testimonial question. Here the famous Williston vignette has its bite:

> In the course of his remarks the Chief Justice stated as one reason for his decision a supposed fact which I knew to be unfounded. I had in front of me a letter that showed his error. Though I have no

25. Frankel, *The Search for Truth—An Umpireal View*, 123 U. PA. L. REV. 1031 (1975).

26. *See* J. WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW, § 1367, at 32 (J. Chadbourn ed. 1974); C. WRIGHT, LAW OF FEDERAL COURTS 398 (3rd ed. 1970).

doubt of the propriety of my behavior in keeping silent, I was somewhat uncomfortable at the time.

One of the troublesome ethical questions which a young trial lawyer is confronted with is the extent to which he is bound to disclose to the court facts which are injurious to his client's case. The answer is not doubtful. The lawyer must decide when he takes a case whether it is a suitable one for him to undertake and after this decision is made, he is not justified in turning against his client by exposing injurious evidence entrusted to him. If that evidence was unknown to him when he took the case, he may sometimes withdraw from it, but while he is engaged as counsel, he is not only not obliged to disclose unfavorable evidence, but it is a violation of his duty to his client if he does so.

In a little book of the historian Lecky, entitled *The Map of Life,* there is a chapter on "Moral Compromise," showing that doing something intrinsically regrettable, because the only alternative involves worse consequences, is a necessity in every profession. . . .That the duty of a lawyer—especially a trial lawyer—should also involve the necessity of choosing the lesser of two evils should therefore not occasion so much surprise as it sometimes does.[27]

The Williston story clearly illustrates the demands of the system. While men like Judge Frankel strive to change the "system,"[28] making the obligation to the "truth" a greater one for each lawyer, at present, the principles of Canon 4 of the Code of Professional Responsibility stoutly proclaim that lawyers are duty-bound to keep a client's confidences and secrets in the circumstances described by Williston.[29]

---

27. S. WILLISTON, LIFE AND LAW 271-72 (1940). I once repeated the Williston story in a speech before a number of lawyers and judges and ended it by saying that, unlike Williston, I thought many lawyers would be very pleased at the result because:

(a) The result was a victory for the side the lawyer was on, and we all like to win;

(b) The result undoubtedly pleased the client and thus the lawyer made another person happy. In addition, the lawyer probably secured payment of a reasonable fee with an increased likelihood of future fees; and

(c) The result would, today, be ethically mandated by Canon 4 of the Code of Professional Responsibility (see note 51 *infra* and accompanying text): to have disclosed the crucial fact would have been cause for disciplinary action against Williston and perhaps even an action in tort against him.

I mention this only because the audience reaction was one I did not expect. There were expressions of outrage at my pronouncements both:

(a) That Williston did what he should have done; and

(b) That many lawyers' reactions would be as described for the reasons described.

I was dismayed. I realized that Freedman was correct when he noted that the implications of the adversary system are not evident to many of those who work within it.

28. *See* Frankel, *From Private Fights Toward Public Justice, supra* note 12; Frankel, *The Search for Truth—An Umpirial View, supra* note 25. See also text accompanying note 25 *supra*.

29. See notes 51-52 *infra* and accompanying text.

It has long been recognized, even by those who have countenanced lying on the part of lawyers in some circumstances, that "it is inadmissible to lie to the court."[30] Indeed, the central positive argument that there is no "right to lie" rests on the fact that perjury and false swearing are typically crimes.[31] In *Harris v. New York*,[32] for example, Mr. Chief Justice Burger noted that: "Every criminal defendant is privileged to testify in his defense, or to refuse to do so. But that privilege cannot be construed to include the right to commit perjury."[33] In *Harris*, the Supreme Court held that a statement, which was inadmissible against the defendant in the prosecution's case in chief because it was obtained in violation of *Miranda*[34] warnings, may still be used to impeach the defendant's credibility should the defendant elect to testify. Four Justices dissented, three of them arguing that *Miranda* would be seriously weakened by such a holding. No dissenting Justice, however, denied the truth of the Chief Justice's statement quoted above.

An additional argument that there is no right to lie is found in the many prohibitions against lying by lawyers contained both in the current Code of Professional Responsibility ("the Code"), as well as in the Canons of Ethics ("the Canons") which preceded the Code.[35] Surely the fact that the lawyer may not lie implies that a

---

30. *Curtis, supra* note 13, at 7-8.

31. At common law they were defined as follows:

Perjury is a false oath in a judicial proceeding in regard to a material matter. A false oath is a willful and corrupt sworn statement made without sincere belief in its truthfulness.

False swearing is what would be perjury except that it is not in a judicial proceeding, but in some other proceeding or matter in which an oath is required by law.

R. PERKINS, CRIMINAL LAW 454 (1969).

32. 401 U.S. 222 (1971).

33. *Id.* at 225.

34. Miranda v. Arizona, 384 U.S. 436 (1966).

35. The Code was adopted on August 12, 1969, to become effective for ABA members on January 1, 1970. It was amended by the House of Delegates in February of 1970, February of 1974, February of 1975, and August of 1977. The primary source of interpretation of the Code is the ABA Ethics Committee Opinions, although state committees, disciplinary boards and judicial bodies also "interpret."

The Code is divided into three parts:

(a) The Canons, which are statements of axiomatic norms;

(b) The Ethical Considerations (ECs), which are aspirational in character; and

(c) The Disciplinary Rules (DRs), which are mandatory. In the words of the Preliminary Statement to the Code: "The Disciplinary Rules state the minimum level of conduct below which no lawyer can fall without being subject to disciplinary action."

ABA CODE OF PROFESSIONAL RESPONSIBILITY, Preamble and Preliminary Statement [hereinafter cited as ABA CODE].

The Code replaced the old Canons, promulgated in 1908 and infrequently amended. The Canons were built largely on Hoffman's *Fifty Resolutions in Regard to Professional Deportment*, in D. HOFFMAN, A COURSE OF LEGAL STUDY 752-75 (2d ed. 1836), and on G.

client or witness may not lie either. This implication is clear from the language of DR 7-102(4) (a lawyer shall not knowingly use perjured testimony or false evidence), DR 7-102(6) (nor participate in the creation or preservation of evidence when he knows or it is obvious that the evidence is false), and DR 7-102(7) (nor counsel or assist his client in conduct that the lawyer knows to be illegal or fraudulent).

Freedman, however, denies the absoluteness of the proposition that there is no right to lie, although his examples seem limited to two situations. The first deals with the criminal defendant on the witness stand. Although placing the argument in the mouths of "a number of experienced attorneys," Freedman declares that "the criminal defendant has a right to 'tell his story.' " By this comment, Freedman suggests that it is "simply too much to expect of a human being, caught up in the criminal process and facing the loss of liberty and the horrors of imprisonment, not to attempt to lie to avoid that penalty."[36] Freedman's second and clearest assertion of a right to lie comes in the context of the criminal defendant's plea. Freedman says that "a criminal defendant is privileged to lie to the court in pleading 'not guilty' even when the defendant knows that the plea is contrary to the fact."[37]

Freedman's the "right to tell his story" and the "not guilty plea" situations are the only examples found that argue the position that a person—other than a lawyer[38]—has a right to lie within the context of the adversary system. These examples are not persuasive. First, whatever one's view about what is or is not "too much to expect of a human being," the adversary system *demands* that the truth be told or that silence be maintained. This demand is one of the central differences between the adversary system of criminal justice and the civil law system. Freedman seems to recognize this in the sentence immediately following the language quoted above. Referring to that language, he goes on to note that "for that reason, criminal defendants in most European countries do not testify under oath, but simply 'tell their stories.'"[39] Precisely. And in the Anglo-American system, the criminal defendant is not required to testify at all; but if he does, he is obliged not to lie.

SHARSWOOD, ESSAY ON PROFESSIONAL ETHICS (5th ed. 1884). The latter evolved from a series of lectures first delivered in 1854. *Id.* at (5). The first set of Canons was adopted by the Alabama State Bar Association in 1887. *See* H. DRINKER, LEGAL ETHICS 23-24 (1953).

36. M. FREEDMAN *supra* note 1, at 31.

37. *Id.*

38. Charles Curtis argues that a lawyer has a duty to lie for his client under certain circumstances. See notes 47-49 *infra* and accompanying text.

39. M. FREEDMAN, *supra* note 1, at 31.

A distinguished Italian legal scholar, commenting on this distinction, praised the approach of the adversary system as a more noble idea, more in keeping with an ideal of human dignity.[40] John Noonan made a similar point in attacking Freedman's position. Although he agreed with Freedman that the importance of preserving human dignity is a fundamental purpose of the criminal trial, Noonan commented that "it seems strange to build on this foundation a defense of lying. 'Only if I can tell my shabby falsehood with your help can I retain my sense of human worth' is what we must take to be the plea of each incriminated defendant who wants to testify falsely."[41]

Freedman's second example is no more convincing, as another refutation by John Noonan makes clear. The not guilty plea, as used in the context of a court proceeding, is not a lie because it is understood by everyone to mean: "I cannot be proven guilty of the charge by the ordinary process of law."[42] More accurately, it is an understood convention used by the defendant to demand his right that the government prove his or her guilt beyond a reasonable doubt. Surely, the criminal "plea" is understood by no one as involving a question of lying, because no one is deceived. It would appear that these unconvincing arguments are advanced primarily to cast doubt on the absolute "no-right-to-lie" principle. By doing so, it becomes easier to argue that a lawyer under some circumstances "may," perhaps "should," either lie for his client or at least assist the client in lying because other principles or values are considered of overriding importance. This topic must now be considered.

2. *No Lawyer Cooperation*—Under our present ethical standards and legal practices, Williston's position in the case previously described cannot be disputed. Nevertheless, to say that a lawyer is duty bound not to disclose facts that may be detrimental to his client's case is not at all to say that a lawyer may assist his client in presenting false evidence or in giving false testimony. The present Code is replete with clear statements on this point. To begin with, there is the broad admonition of DR 1-102(A)(4): "A lawyer shall not: Engage in conduct involving dishonesty, fraud, deceit or misrepresentation." DR 7-102 asserts the bounds of permissible advocacy, and explicitly states that a lawyer shall not "knowingly use

40. Lecture by Professor Giovanni Bognetti, class in Professional Responsibility at the Case Western Reserve School of Law (1975). Professor Bognetti holds the chair in Public Law at the University of Pavia, Milan, Italy.

41. Noonan, *supra* note 2, at 364.

42. Noonan, *The Purposes of Advocacy and the Limits of Confidentiality*, 64 MICH. L. REV. 1485, 1491-92 n.28 (1966).

perjured testimony or false evidence" (A)(4); nor shall he "knowingly make a false statement of law or fact" (A)(5); nor "participate in the creation or preservation of evidence when he knows or it is obvious that the evidence is false" (A)(6); nor "counsel or assist his client in conduct that the lawyer knows to be illegal or fraudulent" (A)(7).

The directives of these disciplinary rules are not new. The Canons reflected a similar ethical stance. Canon 15 stated: "The office of attorney does not permit, much less does it demand of him for a client, violation of law or any manner of fraud or chicane." Canon 22 read: "It is unprofessional and dishonorable to deal other than candidly with the facts in taking the statement of witnesses, in drawing affidavits and other documents, and in presentation of causes." The roots of these principles are deep in history.[43]

Not only is the lawyer prohibited from lying, but he has a positive obligation to make known the perjury or fraud of another. Former Canon 29 provided that: "The counsel upon the trial of a cause in which perjury has been committed owes it to the profession and to the public to bring the matter to the knowledge of the prosecuting authorities." Similarly, DR 7-102(B)(1) as originally adopted,[44] and as it still reads in many state Codes,[45] provides:

> A lawyer who receives information clearly establishing that:
>
> His client has, in the course of the representation, perpetrated a fraud upon a person or tribunal shall promptly call upon his client to rectify the same, and if his client refuses or is unable to do so, he shall reveal the fraud to the affected person or tribunal.[46]

---

43. The lawyer's oath in Massachusetts in 1701 contained these words: "You shall do no falsehood, nor consent to any to be done in the court, and if you know of any to be done you shall give knowledge thereof to the Justices of the Court, or some of them, that it may be reformed." In 1307, the English bar was obliged to follow a similar rule:

> Every pleader is to be charged by oath that he will not maintain nor defend what is wrong or false to his knowledge . . . he [is] to put on before the Court no . . . false evidence, nor move nor offer any corruptions, deceits, tricks or false lies, nor consent to any such. . . .

J. COHEN, THE LAW: BUSINESS OR PROFESSION? 87-88 (1916).

44. ABA CODE, *supra* note 35.

45. As of October 1976, only seven states had adopted the 1974 amendment, discussed at note 46 *infra*. ABA ETHICS COMMITTEE, ANNOTATED CODE OF PROFESSIONAL RESPONSIBILITY 227 (Oct. 1976) (unpublished draft), *cited in* Wolfram, *Client Perjury*, 50 S. CAL. L. REV. 809, 865 n. 218 (1977).

46. The ABA Code was amended in 1974 to add the following clause at the end of DR 7-102(B)(1): "except when the information is protected as a privileged communiation." Important aspects of the controversy surrounding the clause itself and the 1974 amendment are discussed in M. FREEDMAN, *supra* note 1, at 28-29. See notes 129-36 *infra* and accompanying text.

Despite these strong admonitions, some lawyers persist in claiming that lawyers may lie for clients or may cooperate with clients in perpetrating a lie. Charles Curtis, for instance, presents several scenarios where he contends that such action is justified. He suggests, for example, that a lawyer is justified in lying about the facts of his representation of a man who had been blackmailed by some other lawyers. As secretary of the grievance committee of his bar association, Curtis asked the lawyer about his representation of the blackmail victim; the lawyer promptly denied ever having known the man. Subsequently, the lawyer persisted in his denial before the grievance committee and eventually before the court itself in disbarment proceedings. Without giving reasons, Curtis rendered the opinion that the lawyer performed admirably before the grievance committee, although Curtis apparently believed that lying to the court was impermissible. On the latter point, Curtis reasoned that "a lawyer's duty to his client cannot rise higher than its source, which is the court."[47] Presumably, even under the Curtis rationale, if lying is forbidden by the Code of Professional Responsibility, as promulgated or approved by the court, or if lying is forbidden by the rules which permit the lawyer to hold his license to practice law, then lying is *always* forbidden. But Curtis does not deal with that troublesome implication.

Curtis has also discussed the right of a lawyer to lie if the person asking the question has no right to ask it. DR 7-106(C)(3) provides that a lawyer "shall not" assert his personal knowledge of the facts in issue. It would thus be improper for a judge to inquire of a lawyer, during a sentencing proceeding, whether the client defendant has a criminal record. In the ABA opinion on this subject, the Ethics Committee said the lawyer "should ask the court to excuse him from answering the question, and retire from the case," despite the certainty that "this would doubtless put the court on further inquiry as to the truth."[48] Clearly this position rests on the strict prohibition against lying by lawyers.

Curtis also considers the situation where a client is a fugitive from justice. It is not at all clear whether the court or the police have a right to ask about the client's whereabouts.[49] Nevertheless, Curtis contends that the lawyer must lie in response. But why? Refusing to answer seems equally satisfactory, if indeed one determines that

47. *Curtis, supra* note 13, at 7-8.

48. ABA COMMITTEE ON PROFESSIONAL ETHICS, OPINIONS, No. 287 (1967). See notes 123-32 *infra* and accompanying text.

49. The "fugitive" cases have produced a mixed bag of responses from ethics committees. *See* A. KAUFMAN, PROBLEMS IN PROFESSIONAL RESPONSIBILITY 116-18 (1976).

confidentiality forbids disclosure in this case.

Neither of the cases Curtis advances justifies lying. He offers no reason for the positions he adopts. Whatever the reasoning process, behind it is the paramount position which "confidentiality" holds for Curtis within the system. "Confidentiality" is the issue for Freedman too. Freedman, however, does not argue that the lawyer should lie for his client; he only argues that the criminal defense lawyer should not disclose confidences. If that means he must participate fully in the examination of a criminal defendant he "knows" to be lying, well, so be it.[50] His argument is that a lawyer must act as if he both believes his client and in his client. To do otherwise, Freedman maintains, would result in a lack of advocacy that both the system and the Constitution condemn. This position will be examined carefully in the following sections of this article.

II. Confidentiality and the Adversary System

The strongest argument against the thesis of this article is found in the standard embodied in Canon 4 of the Code. Canon 4 sets forth the broad prohibition against lawyers revealing a confidence or secret[51] or using information gained in the professional relationship to the disadvantage of the client.[52] Monroe Freedman uses this protection of "confidentiality" to bolster his position that a criminal lawyer must cooperate with a lying client because to do otherwise would violate Canon 4's prohibitions. Although Freedman makes some additional points, it is clear that "confidentiality" is the lynchpin in his argument insofar as that argument is systemic.[53] Thus, the principle of "confidentiality" must be examined more closely.

The Code prohibition against breaching confidences and secrets is modified by four exceptions. DR 4-101(C) provides that a lawyer may reveal:

(1) Confidences or secrets with the consent of the client or clients affected, but only after a full disclosure to them.

(2) Confidences or secrets when permitted under Disciplinary Rules

---

50. M. Freedman, *supra* note 1, at 40.

51. "Except when permitted under DR 4-101(C), a lawyer shall not knowingly: (1) Reveal a confidence or secret of his client." ABA Code, *supra* note 35.

52. DR 4-101(B). "Except when permitted under DR 4-101(C), a lawyer shall not knowingly: . . . (2) Use a confidence or secret of his client to the disadvantage of his client." *Id.*

DR 4-101(B)(3) goes even further. A lawyer shall not use a confidence or secret for his own advantage or for the advantage of a third person unless his client consents after full disclosure. However, the caveat remains: "Except when permitted under DR 4-101(C)."

53. *See* M. Freedman, *supra* note 1, at 1-8, 27-42.

or required by law or court order.
(3) The intention of his client to commit a crime and the information necessary to prevent the crime.
(4) Confidences or secrets necessary to establish or collect his fee or to defend himself or his employees or associates against an accusation of wrongful conduct.

On the other hand, the prohibitions against "lying" contained in the Code[54] are unequivocal and contain no exceptions. Therefore, to find that "confidence" overrides "lying" requires a strained construction of the Code.[55]

A. Confidentiality: The Policies

Wigmore, in tracing the history of the attorney-client evidentiary privilege, has also identified the justifications given for the ethical principles concerning confidentiality.[56] The two have historically been treated as nearly identical, the ethical principle of "confidentiality" being somewhat broader in scope because it exists without regard to "the nature or source of information or the fact that others share the knowledge."[57] Also, the ethical principle must be broader than the evidentiary privilege in order to cover out-of-court statements. It is clear, however, that the policies behind each are identical.[58]

The attorney-client evidentiary privilege was first based on what was at least an attorney's implied duty to maintain his clients' secrets because of the attorney's oath and honor. Thus it was the lawyer's privilege, not the client's. However, in Wigmore's words, "[t]he judicial search for truth could not endure to be obstructed

54. See notes 35-42 *supra* and accompanying text.

55. Some attempts at such constructions have been made by ABA Ethics Committees. *See* Formal Opinion No. 287, *supra* note 48; ABA COMMITTEE ON PROFESSIONAL ETHICS, OPINIONS, No. 341 (1975). The tortured nature of these opinions will be discussed at notes 120-40 *infra* and accompanying text.

56. J. WIGMORE, EVIDENCE §§ 2290-2291, at 542-54 (McNaughton rev. 1961).

57. ABA CODE, *supra* note 35, EC 4-4. The present Code seems to be the first place where the distinction between confidentiality and the evidentiary privilege was clearly made. However, the footnotes to the ECs and DRs of Canon 4 demonstrate that the foundations and policies of both are identical. Under the Code, the word "confidence" refers to information that is "protected by the attorney-client privilege;" "secret" refers to other information gained in the professional relationship that the client requests be withheld from disclosure or that, if disclosed, may be embarrassing or detrimental to the client. ABA CODE, *supra* note 35, DR 4-101(A).

58. ABA CODE, *supra* note 35, DR 4-101(A). Under the Canons, the discussion of the application of the confidentiality rule was considered largely a question of law, not ethics. H. DRINKER, *supra* note 35, at 132. A two-volume treatise on attorneys, published less than 40 years before Drinker's book, did not consider the two separate. E. THORNTON, ON ATTORNEYS AT LAW 154-71 (1914).

by a voluntary pledge of secrecy."[59] By the last quarter of the eighteenth century, the doctrine was entirely repudiated. It was replaced, however, with a new theory, one which was to gain a secure foothold in Anglo-American jurisprudence. The new theory was that a privilege against disclosure by a lawyer of confidential information was necessary to insure that the client would freely communicate all details of the matter of the retainer without fear that some or all of the information would be disclosed by the lawyer to the client's harm. This new privilege was the client's. Moreover, that policy remains the bulwark of the evidentiary privilege and the broader ethical principle of confidentiality found in Canon 4 of the Code.[60]

The attorney-client privilege was never without its critics, the most outstanding of whom was Jeremy Bentham. In a pithy summary of the critic's argument, Wigmore explained that for Bentham, "[i]t always comes back to this, that the deterring of a guilty man from seeking legal advice is no harm to justice, while the innocent man has nothing to fear and therefore will not be deterred."[61] Wigmore answered Bentham with a variety of arguments suggesting principally that matters are never so clearly black or white as Bentham implied. Moreover, he suggested that "the sense of treachery in [a counselor] disclosing such confidences is impalpable and somewhat speculative, but it is there nevertheless."[62] Bentham sarcastically dismissed the contention that an attorney's honor as a person of moral worth was also involved in this topic. Without the privilege, as Wigmore says, "the position of the legal advisor would be a difficult and disagreeable one, for it must be repugnant to any honorable man to feel that the confidences which his relation naturally invites are liable at the opponent's behest to be laid open through his own testimony."[63] Wigmore recognized that a natural feeling of betrayal would arise within any client who confided in his attorney only to discover later that his attorney had disclosed those confidences to the client's harm. The word "treacherous" is not too strong to characterize the behavior of such an attorney. This question of "treachery" will be discussed later in this article,[64] because of its persistent emotional strength. Although the substantive law of evidence has rejected the old honor-treachery grounds for support

59. J. WIGMORE, *supra* note 56, § 2290, at 543.

60. *Id.* § 2291, at 543. *See also* M. FREEDMAN, *supra* note 1, at 4-5.

61. J. WIGMORE, *supra* note 56, § 2291, at 552. Though the use of the words "guilty" and "innocent" do not apply to civil litigation, Wigmore clearly focuses his discussion on civil cases too.

62. *Id.* at 553.

63. *Id.*

64. See notes 152-54 *infra* and accompanying text.

of the attorney-client privilege, lawyers nevertheless continue to experience emotional concern when faced with the prospect of revealing client confidences.

Wigmore concludes this discussion of the attorney-client privilege with this important observation:

> Nevertheless, the privilege remains an exception to the general duty to disclose. Its benefits are all indirect and speculative; its obstruction is plain and concrete. Even the answers to Bentham's argument concede that the privilege is well founded in its application to a certain proportion of cases. It is worth preserving for the sake of a general policy, but it is nonetheless an obstacle to the investigation of the truth. It ought to be strictly confined within the narrowest possible limits consistent with the logic of its principle.[65]

Wigmore suggests a narrow construction of the evidentiary privilege that is, of course, itself narrowly defined to begin with.[66] Thus it should not be surprising to find that the ethical principle of confidentiality, broader in definition than the attorney-client privilege and yet resting on identical policy foundations, has exceptions to it that are extremely broad.[67]

Freedman, in discussing the place of confidentiality in the system, ignores entirely the history traced by Wigmore. By doing so, Freedman places undue emphasis on the principle of confidentiality because it better comports with his idea of the system. The adversary system, however, is an historical reality as well as a philosophical system of adjudication. To ignore history is simply to deny its continuing impact upon the present. Holmes taught us this so long ago that his maxim is a motto for almost everything we say about law: "The life of the law has not been logic: it has been experience."[68] Moreover, after extensively quoting a number of sources to establish the current policy ground of allaying clients' fears about disclosure, Freedman states that:

> Accordingly, the new Code of Professional Responsibility provides that a lawyer shall not knowingly reveal a confidence or secret of the client, nor use a confidence or secret to the disadvantage of the client, or to the advantage of a third person, *without the client's consent.*"[69]

---

65. J. WIGMORE, *supra* note 56, § 2291, at 554 (citations omitted).

66. *See id.* Wigmore notes that the privilege applies only: (a) where legal advice of any kind is sought, (b) from a professional legal advisor, (c) the communications relating to that purpose, (d) made in confidence, (e) by the client, (f) are at his instance permanently protected, (g) from disclosure by himself or by the legal advisor, (h) except the protection be waived.

67. See notes 51-55, *supra* and accompanying text.

68. O. HOLMES, THE COMMON LAW 5 (1881).

69. M. FREEDMAN, *supra* note 1, at 5 (emphasis added).

He stops there. He does not indicate that besides the situation where the client consents, there are three additional broad exceptions in the Code.[70] This less than candid description of the Code's position on confidentiality seriously cripples Freedman's argument that "confidentiality" has the central place in our system that he claims for it.[71]

In fairness to Freedman, he does raise constitutional considerations, which must be discussed in dealing with the issue of "confidentiality."[72] To a great extent, however, the Constitution simply presupposes the adversary system or perhaps incorporates it. In any event, for analytical purposes, constitutional matters are kept distinct from systemic matters in this article and Freedman's constitutional concerns will be discussed later. Since they are limited to criminal cases, it is more appropriate to discuss them in that context.

B. A Paradigm Case

In order to provide concreteness and a necessary base for the comparison of other cases, confidentiality will now be discussed within the context of a paradigm case. The case is the extremely difficult one Monroe Freedman uses as his focus in discussing lawyers' ethics in a case of perjury:

> Assume the following situation. Your client has been falsely accused of a robbery committed at 15th and P Streets at 11:00 P.M. He tells you at first that at no time on the evening of the crime was he within six blocks of that location. However, you are able to persuade him that he must tell you the truth and that doing so will in no way prejudice him. He then reveals to you that he was at 15th and P Streets at 10:55 that evening, but that he was walking east, away from the scene of the crime, and that, by 11:00 he was six blocks away. At the trial, there are two prosecution witnesses. The first mistakenly, but with some degree of persuasiveness, identifies your client as the criminal. At that point, the prosecution's case depends upon the single witness, who might or might not be believed. The second prosecution witness is an elderly woman who is somewhat nervous and who wears glasses. She testifies truthfully and accurately that she saw your client at 15th and P Streets at 10:55 P.M. She has

70. ABA CODE, *supra* note 35, DR 4-101(C)(2)-(4). See notes 53-54 *supra* and accompanying text.

71. Freedman does mention the "future crime" exception in passing, but does not clearly indicate in his text that he recognizes the breadth of the Code's specific exceptions. *See* M. FREEDMAN, *supra* note 1, at 6.

72. These constitutional considerations are discussed at notes 108-119 *infra* and accompanying text.

> corroborated the erroneous testimony of the first witness and made conviction extremely likely. However, on cross-examination her reliability is thrown into doubt through demonstration that she is easily confused and has poor eyesight. Thus the corroboration has been eliminated, and doubt has been established in the minds of the jurors as to the prosecution's case.
>
> The client then insists upon taking the stand in his own defense, not only to deny the erroneous evidence identifying him as the criminal, but also to deny the truthful, but highly damaging, testimony of the corroborating witness who placed him one block away from the intersection five minutes prior to the crime. Of course, if he tells the truth and thus verifies the corroborating witness, the jury will be more inclined to accept the inaccurate testimony of the principal witness, who specifically identified him as the criminal.[73]

1. *Freedman's Solution*—Freedman answers the ethical question posed by the paradigm in this way:

> In my opinion, the attorney's obligation in such a situation would be to advise the client that the proposed testimony is unlawful, but to *proceed in the normal fashion* in presenting the testimony and arguing the case to the jury if the client makes the decision to go forward. Any other course would be a betrayal of the assurances of confidentiality given by the attorney in order to induce the client to reveal everything, however damaging it might appear.[74]

The Code clearly states that "[i]n his representation of a client, a lawyer shall not . . . *knowingly* use perjured testimony or false evidence;"[75] nor shall a lawyer "counsel or assist his client in conduct that the lawyer *knows* to be illegal or fraudulent."[76] Nevertheless, Freedman asserts that criminal lawyers, at least in Washington, D.C., regularly do as he suggests.[77] Recognizing this discrepancy, Freedman suggests that the Code may not really mean what it appears to say. He suggests, therefore, that we either (a) look at the cases cited by the Code draftsmen to try to devise an interpretation of these disciplinary rules which will accurately "gloss" the Code, or (b) avoid the ethical problem by reading the words

---

73. M. FREEDMAN, *supra* note 1, at 30-31.

74. *Id.* at 31 (emphasis added).

75. ABA CODE, *supra* note 35, DR 7-102(A)(4) (emphasis added).

76. *Id.* DR 7-102(A)(7) (emphasis added).

77. M. FREEDMAN, *supra* note 1, at 38-39. Freedman reports that ninety percent of the lawyers responding to a survey on the question posed by the paradigm case said they would behave exactly as Freedman suggests they should. *Id.* at 38. The information on the survey came from *Freedman, Professional Responsibility in D.C.: A Survey,* 1972 RES IPSA LOQUITUR 60. One commentator has labelled this survey "rather unscientific." *See* Wolfram, *supra* note 45, at 818 (1977).

"knowingly" and "knows" as if they had no content, because to do otherwise would be to compromise a lawyer's role as an advocate in an adversary system.

(a) Cases cited in footnotes to Canon 7—Of the three key cases cited in the footnotes to Canon 7, Freedman summarily disposes of two of them. He argues that the strongest case against his position on confidentiality, *In re Carroll,*[78] is cited as a footnote to an ethical consideration and is thus "aspirational." Freedman found the second case, *Hinds v. State Bar,*[79] to be distinguishable since it involved the lawyer's participation in the fraud itself. Moreover, Freedman found both of these cases to be beside the point because they were civil, not criminal, in nature.

Freedman chose to rely on a third case, *Johns v. Smyth,*[80] as authority for his position. In that case, the court determined that a criminal defendant was entitled to a new trial because of the course followed by his attorney. The allegation that counsel had acted incompetently was based on several points. First, the accused did not testify. Second, proper instructions were not submitted to the trial judge on behalf of the defendant, although it was possible for the defendant to have been convicted of the lesser charge of involuntary manslaughter. The court's instructions failed to mention the possibility of a manslaughter verdict. Finally, counsel had agreed with the prosecutor that the case would be submitted to the jury without argument of counsel. The court noted that, taken alone, these factors would not point to any irregularity as each might properly be considered a trial tactic.[81] Defendant's attorney, however, testified that he failed to argue the case to the jury because he could not do so in good conscience. The attorney doubted the truth of a statement the defendant made to the prosecutor, which statement the prosecutor used as "evidence."

---

78. 244 S.W.2d 474 (Ky. 1951). The citation to the *Carroll* case is found in note 45 to EC 7-26 of the Code. The language from the case that the Code's draftsmen chose to quote bears repeating:

> Under any standard of proper ethical conduct an attorney should not sit by silently and permit his client to commit what may have been perjury, and which certainly would mislead the court and the opposing party on a matter vital to the issue under consideration. . . .
>
> . . .
>
> Respondent next urges that it was his duty to observe the utmost good faith toward his client, and therefore he could not divulge any confidential information. This duty to the client of course does not extend to the point of authorizing collaboration with him in the commission of fraud.

244 S.W.2d at 474-75.

79. 19 Cal. 2d 87, 119 P.2d 134 (1941).

80. 176 F. Supp. 949 (E.D. Va. 1959).

81. *Id.* at 952.

The court's opinion made clear that it was the attorney's duty to use his best efforts to aid his client when the prosecution used the statement in furtherance of its case. The court further observed that such failure to argue the case before the jury "manifestly enters the field of incompetency when the reason assigned is the attorney's conscience."[82] In summarizing his remarks on the case, Judge Hoffman said that "the entire trial in the state court had the earmarks of an *ex parte* proceeding."[83]

Freedman's efforts to construct from the footnote cases a logical rule to supplant the clear expression of intention by the draftsmen of the Code is doomed to failure from the start. The Code's footnotes were never intended to be an annotation of the draftsmen's meaning. This is clearly stated in the first footnote to the Preamble of the Code.[84]

Beyond this defect, the weakness of Freedman's argument is demonstrated in his treatment of the footnote cases. The courts in *Carroll* and *Hinds* condemned the lawyers in each case for proceeding in the "normal fashion" when their clients perpetrated frauds on the court. Both lawyers were told they had a duty to disclose the fraudulent acts to the court. Both were suspended from practice for failing to do so. Freedman's dismissal of these two cases because they were "civil" rather than "criminal" is totally unjustified; neither the Code nor the cases suggest different ethical standards because of the nature of the case at bar. Finally, Freedman discounts the importance of *Carroll* because it appears as a footnote to an ethical consideration, yet wholly embraces *Johns v. Smyth,* which likewise was a footnote to an ethical consideration. Why Freedman claims that the placement of the latter footnote is not significant, while that of the former is significant, is difficult to understand.

Moreover, Freedman's assertion that *Johns v. Smyth* "held that a defendant's constitutional rights had been violated because the attorney, believing his client to be guilty, did not argue the case in the ordinary manner," is both misleading as a statement in the context of his discussion and irrelevant as proof of Freedman's point. As the foregoing discussion of the case suggests, *Johns v. Smyth* held that a criminal defendant is denied effective repesenta-

82. *Id.* at 953.

83. *Id.*

84. The footnotes are intended merely to enable the reader to relate the provisions of this Code to the ABA Canons of Professional Ethics . . . , the Opinions of the ABA Committee on Professional Ethics, and a limited number of other sources; they are not intended to be an annotation of the views taken by the ABA Special Committee on Evaluation of Ethical Standards.

ABA CODE, *supra* note 35, Preamble and Preliminary Statement, n.1.

tion when his counsel refuses to do *anything* for him during the course of the trial *because* counsel believed him guilty of first degree murder and wanted to do nothing that would hinder a conviction. If he had done *nothing* because he thought the tactics were best for his client, or if he had at least argued "reasonable doubt" to the jury, one could not be at all sure that the judge would have held the same way. Thus, Freedman's statement of the holding is misleading. His phrase "did not argue the case in the ordinary manner" implies that the mistrial was called because the lawyer failed to argue perjured testimony to the jury. Clearly, the case does not hold that. The lawyer failed to do *anything* for his client, and thus misunderstood the adversary system as much as Freedman does. "The duty of a lawyer, both to his client and to the legal system, is to represent his client zealously within the bounds of the law."[85]

A broader examination of the case law reveals no support for Freedman's solution to the paradigm case. In a case similar to the paradigm, *McKissick v. United States,*[86] the lawyer for a criminal defendant discovered that his client had committed perjury during an earlier stage of the proceedings. The client told his lawyer of the perjury after the fact. After this disclosure, the lawyer told the judge and requested a mistrial. In ordering the mistrial, the judge noted that it would have been wrong for counsel to have proceeded in the normal fashion:

> The statement [of the client to the lawyer admitting the perjury] was good cause to the attorney to withdraw from the case, and he would have been subject to discipline had he continued in the defense without making a report to the court. The attorney not only could, but was obligated to make such disclosure to the court as necessary to withdraw the perjured testimony from the consideration of the jury. This was essential for good judicial administration and to protect the public.[87]

In re Carroll and other civil cases describe additional instances where discipline (even disbarment)[88] has resulted where lawyers continued with "ordinary" representation after their clients had committed perjury. Thus neither the Code nor the cases substantiate Freedman's position insofar as he attempts to support that position by "systemic" arguments.

---

85. *Id.* EC 7-1.

86. 379 F.2d 754 (5th Cir. 1967).

87. *Id.* at 761.

88. *See In re* Hardenbrook, 135 App. Div. 634, 121 N.Y.S. 250, *aff'd mem.* 199 N.Y. 539, 92 N.E. 1086 (1910), *appeal denied,* 144 App. Div. 928, 129 N.Y.S. 1126 (1911) (disbarred); *In re King,* 7 Utah 2d 258, 322 P.2d 1095 (1958).

(b) *The lawyer's role as advocate*—Freedman notes that there are attorneys who contend that a lawyer can never "know" whether a client is guilty or whether he is lying. For these lawyers, the Disciplinary Rules proscribing the "knowing" use of perjured testimony are therefore irrelevant. Unlike this group, however, Freedman acknowledges the clarity of the Code's mandates. He protests, however, "that the Code does not indicate how the lawyer is to go about fulfilling this obligation."[89]

It is difficult to understand how, after recognizing such a clear obligation, Freedman can advise a course of action in direct contravention of it, simply because "no one told me how to fulfill it." Freedman, as noted previously, ignores the fact that DR 4-101(C)(2) allows the disclosure of a confidence when "permitted under Disciplinary Rules." Some Disciplinary Rules not only "permit," but require[90] conduct which may well entail the disclosure of a confidence. The Code's statement of the principle of confidentiality contains exceptions which permit or require disclosures of confidences in order to prevent a lawyer's participation in the corruption of the system which lying engenders.

Because the criminal bar of the District of Columbia or some members of it may be hypocritical or unethical, or both, does not constitute a valid reason for ignoring the clear mandate of the Code. The fact that this or any group of lawyers may believe that the adversary system requires behavior that exalts client confidentiality beyond every other principle of the system is an insufficient and woefully muddled reason for enshrining that belief as an ethical norm—particularly when it is contrary to otherwise clear ethical rules. Again, the issue is focused by a cool appraisal of the adversary system and a lawyer's role within it.

The adversary system allows trained professionals to assist lay people in presenting facts and arguing the legal significance of those facts. Simply put, it is a system designed to help untrained people accomplish legally permissible aims. But that is all that it is. When lying is condoned through a mistaken assumption about the place "confidentiality" has in the system, the system itself becomes endangered. Moreover, what damage would be done to the principle

89. M. FREEDMAN, *supra* note 1, at 21-24.

90. Although DR 4-101(C)(2) is introduced by the permissive "may," this simply means that confidentiality is no longer a possible argument to inhibit actions that are mandated by other Code provisions. For example, a lawyer must not perform any of the actions forbidden under, *e.g.*, DR 7-102(A), 7-109(A), or 1-102, and *must* perform the obligations required by provisions such as DR 7-102(B) or 7-106(B). See also notes 133-35, *infra* and accompanying text.

of "confidentiality" if a lawyer refused to cooperate with a lying client? The answer to this question probably does depend in part on what the lawyer does to divorce himself from the lie.[91] That issue aside for a moment, the inquiry focuses on three questions: (1) would clients cease coming to lawyers, (2) if they still came, would they be less candid, and (3) even if an appreciable number of clients were to cease coming or if candidness were to suffer in an appreciable number of cases, are these results which we might accept, given the benefits to be derived from making absolute the principle that a lawyer may never cooperate with a client's lie?

It would be helpful in determining the answers to questions (1) and (2) to have some empirical data. Absent that, however, there are presently enough exceptions to the confidentiality principle[92] that the clarification of one situation involving perjured testimony will not create widespread changes in client behavior. Even if we were convinced, however, that client behavior would change if a clear rule were established requiring the disclosure of confidential information by the lawyer in perjury cases, it would still be a debatable question whether the positive value of increased client trust would be worth the negative results. These results may include not only a loss of "truth," but an increase in lawyer cooperation with all manner of questionable client activities under the absolutism of "confidentiality."

Most lawyers would probably agree with the argument presented herein, if it were not for the problems created by (1) the other proposed solutions to the paradigm; (2) the constitutional issues raised by Freedman in the criminal area; and (3) the confusion created by the handling of the "past fraud" issue by ABA ethics opinions. I shall conclude this article by dealing with these three matters seriatim. In dealing with my own conclusions concerning the question discussed in this article, I shall also deal with the question of "treachery," which I believe has an emotional part to play in this area that transcends all of the issues discussed and arguments advanced. The "image" one holds of the adversary system and a lawyer's role in it, I shall contend, makes all the difference.

2. *Other Solutions to the Paradigm*—Before examining other solutions to Monroe Freedman's paradigm case, it may be useful to refocus the case. Freedman's paradigm clearly "stacks the deck" to

91. The proposed "solutions" to the paradigm will be discussed at notes 93-107 *infra* and accompanying text.

92. See notes 53-54 *supra* and accompanying text.

insure the reader's sympathy. His case involves a client, "falsely accused of a robbery." One witness "mistakenly" identifies the client as a criminal. The second witness accurately identifies the client as being near the scene of the crime. The client merely wants to deny the accuracy of the second identification. Our hearts go out to this poor, innocent defendant, and we are softened and made ready for the Freedman solution. But what if the client were not innocent? What if the robbery involved a gun or someone was seriously injured during the commission of the crime? In terms of probabilities, these alterations of the paradigm case seem more likely than Freedman's rather odd set of circumstances. Changing the facts of the paradigm is not meant to be cute or to suggest that innocent people do not become involved in criminal trials. The point is, simply, that it should make no difference whether the client is guilty or innocent. The issue is: What does the adversary system demand of a lawyer? And the answer is: The lawyer must do all he can to represent the client "zealously *within the bounds of the law.*"[93] This is merely another instance of a good end not justifying a bad means. If lying is abhorrent to the system, it is not less abhorrent because it is done to try to insure that a person is not found guilty of a crime he did not commit. Though there is disagreement on this point,[94] it seems to me that the adversary system demands primary allegiance to processes and to institutions.[95] Otherwise, no moral lawyer could defend a guilty defendant, at least not when the lawyer was convinced the defendant was a dangerous, anti-social person. With this as preface, some of the proposed solutions to the Freedman paradigm may be examined.

In a recent article, Dan Aaron Polster argues that the solution to the client fraud dilemma in the paradigm case should be as follows:

---

93. ABA CODE, *supra* note 35, EC 7-1 (emphasis added).

94. Professor Robert Keeton has said:

> The Code of Professional Responsibility does not refer to the use in general of surprise tactics, but clearly the use of such tactics to defeat an admittedly just claim or defense is not supportable. On the other hand, the use of surprise to expose falsification is clearly justifiable. The intermediate ground presents the timeless controversy of "means and ends."

R. KEETON, TRIAL TACTICS AND METHODS § 4-5 (2d ed. 1973). My answer to this is no! The end is a fair trial, not a substantive result. The end is skilled advocacy for a client, not different advocacy, depending on the lawyer's belief in right or wrong in the case at bar. In sum, advocacy has its limits. We must find out what they are.

95. This is Professor Lon Fuller's point. "The Lawyer's highest loyalty is at the same time the most intangible. It is a loyalty that runs, not to persons, but to procedures and institutions." *ABA Report, supra* note 10, at 1162.

> The lawyer would instruct a client who insisted upon taking the stand and committing perjury that, should the client actually begin to testify falsely, the lawyer would approach the bench and inform the judge that his client had just committed perjury. The lawyer would also instruct the client that he (the lawyer) would be a chief witness in the client's subsequent perjury trial.[96]

Polster claims that his position, among all others suggested, best respects the integrity and autonomy, of both the client and the lawyer.[97] What must be undermined is the principle of confidentiality, and thus the attorney-client relationship, though it should be added that Polster's position would require the lawyer in advance to explain to the client the ethical limits to his advocacy.[98] Candor to the client is a basic norm for every lawyer, though of course there are practical problems, particularly when the client is a mistrustful indigent who believes that the lawyer and the prosecutor are in cahoots. For a lawyer to tell that kind of client that, under some circumstances, his "lies" will be revealed to the court may mean that the lawyer will be ineffective because he will never be trusted. Still, it is impossible to deny the essential "rightness" of candor with all clients.

Another solution is embodied in Section 7.7 of the ABA Standards Relating to the Prosecution Function and the Defense Function ("ABA Standards").[99] This solution adopts a principle espoused by Mr. Chief Justice Burger[100] and by Gerald Gold,[101] a prominent criminal lawyer. Under this standard, the lawyer should first attempt to withdraw. Neither Freedman nor Polster accept this as a viable alternative, because it amounts to buck-passing. As Polster puts it, one of three things will happen if a withdrawal takes place: (1) the client will find an attorney who accepts Freedman's theory of the advocate's duty; or (2) the client will lie to his next attorney so he will be free to lie on the stand; or (3) a series of mistrials will be had as ethical lawyer after ethical lawyer refuses

---

96. Polster, *The Dilemma of the Perjurious Defendant: Resolution, Not Avoidance*, 28 CASE W. RES. U.L. REV. 3, 34 (1977).

97. *Id.* at 38.

98. *Id.* at 34-36.

99. ABA PROJECT ON STANDARDS FOR CRIMINAL JUSTICE, STANDARDS RELATING TO THE PROSECUTING FUNCTION AND THE DEFENSE FUNCTION, (approved draft 1971). Although the standards are not widely adopted, they are being used by courts as guides and are widely cited. *See* A. KAUFMAN, *supra* note 49, at 217.

100. Burger, *Standards of Conduct for Prosecution and Defense Personnel*, 5 AM. CRIM. L.Q. 11, 12-13 (1966). At the time of the writing, the Chief Justice was a judge in the D.C. Circuit.

101. Gold, *Split Loyalty: An Ethical Problem for the Criminal Defense Lawyer*, 14 CLEV.-MAR. L. REV. 65, 68-71 (1965).

to cooperate with the lie.[102] It is not at all clear, however, that those three alternatives exhaust the possibilities. Confronted with a lawyer or lawyers unwilling to cooperate in a lie, the defendant may well change his mind.[103] If this suggestion seems naive, it may be considered more "realistic" if one reflects on similar situations where clients might want to do other things clearly beyond the law: fabricate documentary evidence, for example, or parade a series of friendly witnesses to corroborate a fabricated story. How does a lawyer react to those suggestions?

In any event, the ABA Standards go on to tell the lawyer what to do where he cannot withdraw and where the client insists on testifying falsely or perhaps surprises the lawyer with false testimony in the midst of his direct examination. In those cases, the lawyer may not conduct the examination "in the conventional manner," and must not participate in the examination nor argue the false facts to the jury as worthy of belief, nor recite or rely on the false testimony in his closing argument.[104]

A hybrid solution of the approaches of Polster and that of the ABA Standards is that of ABA Informal Opinion Number 1314.[105] That opinion suggests that if the attorney knows of the intended false testimony beforehand, he must either withdraw or tell the court of the falsity. If the lawyer finds out in the middle of the trial that the client has lied, he may "withdraw" without disclosure to the court if his client refuses or is unable to rectify the fraud. The opinion does not say what he is to do if the court refuses to allow him to withdraw. Moreover, the opinion does not indicate why it is correct to tell the court beforehand of the intended falsity but not tell of it after the fact. The committee opinion writer probably knew

---

102. Polster, *supra* note 96, at 34.

103. Although the lawyer did not resign, the client-defendant did change his mind upon his attorney's threatened resignation in United States v. Gonzales, 435 F.2d 1004 (10th Cir. 1970).

104. ABA STANDARDS, *supra* note 99, at § 7.7. Interestingly, if we are to believe the draftsmen of the ABA Standards and the "experienced" criminal lawyers who were consulted on the problem, the situation rarely occurs, at least in private practice. These sources suggest that paying clients rarely admit facts to their lawyers and then insist on testifying falsely. Private attorneys also are believed to have greater leverage with their clients. Presumably, then, this may leave the matter in the laps of criminal lawyers called upon to defend indigents. *Id.* at 276. Absent "harder" data, however, I remain skeptical about these behavioral assertions.

105. ABA COMMITTEE ON ETHICS AND PROFESSIONAL RESPONSIBILITY, INFORMAL OPINION No. 1314 (1975). An informal opinion is supposed to deal with a comparatively narrow question which arises infrequently. ABA COMMITTEE ON PROFESSIONAL ETHICS, *supra* note 48, at 6. Since informal opinions are usually brief and contain little analysis and few citations, their value as authority is slight. This opinion is discussed in the text only as evidence of another possible solution to the paradigm.

that it was permissible under the Code to tell of an intention to commit a crime, while the propriety of disclosing a past fraud is doubtful when the disclosure would violate a privileged communcation.[106] Rather than come to terms with underlying policies or principles, the opinion tried to be legalistic and failed to see how "unrealistic" its answer was. If the lawyer tried to withdraw in the middle of the case, the court would not normally grant the request without a good reason. Disclosure would almost inevitably follow.

Of course, the ABA Standards are equally unrealistic as Polster and Freedman point out. Refusing to examine the client and to argue the case either allows the "lie" to slip into the jury box or else demonstrates to the trier of fact that the lawyer disbelieves his client. The first result is totally unacceptable because no lie is permitted under the system, nor may a lawyer allow the system to act on the basis of the lie. The counter-argument, of course, is that the lawyer did not "use" perjured testimony or "assist" the client in his lie. Although this may be technically correct, these gestures of nonparticipation and silence will be interpreted by the trier of fact. Either they will be understood to mean that the lawyer countenances these statements, in which case the lawyer violates a fundamental principle of the system, or else his gestures will be seen as saying that he disbelieves his client, in which case the lawyer's idea of "confidentiality" is unrealistic in the extreme.

The Polster position may go too far, however, in having the lawyer testify at the trial of the client on the perjury charge. This seems to be an altogether new suggestion. It is one thing to ask that a lawyer ensure that "lying" not be condoned, and quite another to force the lawyer to be the enforcer in a criminal context against a client who has confided in him. This first case says: "Don't use me for your corrupt practices." The second says: "If you do, I will harm you in a positive way." It smacks of retribution and retaliation.

Although variants are no doubt possible, the Freedman, Polster,[107] ABA Standards and Informal Opinion 1314 positions seem to run the gamut of the possibilities. Freedman and Polster represent the extremes: the former arguing for strict "confidentiality," the latter arguing for forthright disclosure and the lawyer's testimony against the client in a subsequent perjury trial. The ABA Standards decree no "active" participation, but no disclosure either. The Stan-

106. *Compare* ABA CODE, *supra* note 35, DR 4-101(C)(3) *with id.* DR 7-102(B)(1). This matter is discussed at notes 122-40 *infra* and accompanying text.

107. A variant of the Polster position may be found in Wolfram, *supra* note 45. Professor Wolfram's article is very comprehensive and discusses the issue in both the civil and the criminal contexts.

dards walk a tightrope that exalts form over substance, producing an unrealistic result. The Informal Opinion is legalistic, inconsistent, and unprincipled. Before attempting to tie together many of the ideas presented in this article into a proposal of my own, the "constitutional" and "past fraud" obstacles must still be cleared away.

III. The Constitutional Confusion

What if the client insists upon taking the stand despite the lawyer's efforts to dissuade him? What if surprises occur, and the lie is "out" before the lawyer can act? Is Freedman's solution the one demanded by the Constitution? Briefly, the answer is "no." And Freedman is alone in suggesting that it is otherwise. Aside from *Johns v. Smyth,* discussed earlier,[108] which does not hold what Freedman implies that it does, Freedman's constitutional arguments largely rest on vague generalities. He states that "divulgence by the defense attorney in a criminal case would be controlled by such constitutional provisions as the right to counsel, the privilege against self-incrimination, the right to trial by jury, and the right to due process."[109]

Freedman does cite one other case, *Holmes v. United States,*[110] for the proposition that a defendant has been denied the right to counsel or of due process when the judge had been informed by defendant's lawyer that the defendant perjured himself. In reality, *Holmes* had nothing to do with the issue Freedman discusses. The majority of the court did remand the case on due process grounds, but the due process issue was solely whether the defendant had actually waived a preliminary hearing. In passing on that issue, the dissenting judge, on whose opinion Freedman focuses, recited the fact that he thought no "prejudice" had occurred. Presumably the judge reached this conclusion because the defendant both lied to his lawyer and on the witness stand during the trial. On remand, the district court found that an effective preliminary hearing had not been denied nor had the defendant been prejudiced. The conviction was therefore affirmed.[111] It is hard to tell what Freedman's point was in citing *Holmes,* but in the context of his argument about constitutional rights it lends no support. In fact, it is *contra* authority. In *United States v. Gonzales,* where a similar fact situation

108. See notes 80-85 *supra* and accompanying text.
109. M. Freedman, *supra* note 1, at 29.
110. 370 F.2d 209 (D.C. Cir. 1966). *See* M. Freedman, *supra* note 1, at 34.
111. 370 F.2d at 213.

occurred,[112] the court dealt head-on with the issue, finding no abuse by the judge nor any denial of the right to counsel.

There are some cases which hold that it is proper for a mistrial to be called when counsel informs the court of client perjury. In *McKissick v. United States,*[113] for example, after learning of his client's perjury, the defendant's lawyer informed the court and asked leave to withdraw. Leave was granted, but in passing on the question of whether the double jeopardy provision of the fifth amendment would be violated by another trial, the court said: "[T]he appellant had no constitutional right, overriding the public interest, to have his case determined by a tribunal whose processes he had himself thus frustrated and abused."[114] The court thus concluded that the district court had not abused its discretion by finding a manifest justification or necessity for declaring a mistrial and had correctly concluded that the double jeopardy provision would not apply. The Fifth Circuit also held, however, that though the trial attorney acted correctly in informing the judge of the client's perjury,[115] the lawyer should not have moved for a mistrial without the client's consent. If no consent was given, the defendant had been denied the effective assistance of counsel in the case. Presumably, if the court had declared a mistrial on its own there would have been no question regarding the effective assistance of counsel.[116] However artfully or inartfully the court drew its distinctions, it was untroubled by Freedman's concerns. Thus the opinion noted that the district court could have directed the lawyer to continue the representation of the defendant even after the counsel disclosed the perjury.[117] In short, the constitutional considerations are beside the point. They represent another Freedman effort to buttress his real concern, confidentiality in the adversary system.

Since the constitutional considerations have no substance, the civil and criminal processes should be treated in the same way for purposes of understanding a lawyer's ethics under the adversary system. Even Freedman concedes that the consensus of the bar on civil matters seems to be against confidentiality in the "lying

112. 435 F.2d at 1008-10.

113. 379 F.2d 754 (5th Cir. 1967). See notes 88-89 *supra* and accompanying text.

114. 379 F.2d at 761.

115. The court noted that had the attorney acted otherwise, he would have been subject to discipline. *Id.*

116. There were other aspects of the case that were important to the decision, such as the factual question of whether the defendant actually admitted the perjury and whether the defendant had a right to confront his lawyer or others in chambers during the discussion of the "perjury" issue. *Id.* at 760-62. These aspects are not relevant to the discussion in the text.

117. *Id.* at 762.

client" cases.[118] The Code itself makes no distinctions between civil and criminal cases. This is true, despite a strange statement in the ABA Standards that the original version of DR 7-102(B)(1) of the Code (mandating the lawyer's duty to disclose his client's past fraud if the client refuses to do so) "is construed as not embracing the giving of false testimony in a criminal case."[119] The Code does, however, single out "past fraud" cases for special treatment. Since the Ethics Opinions and the provisions of the old Canons and the Code in the "past fraud" area are so crucial to the question of confidentiality and the adversary system, these will now be considered.

IV. Past Fraud and the ABA

Before discussing the relevant ABA documents, some terminology needs to be clarified. In dealing with the issue of "fraud," three kinds of cases should be distinguished. They may be labeled "past," "present," and "future." The case of "past fraud" deals with events of the past that have an air of finality to them, as where the lawsuit has terminated; the case of "present fraud" deals with events that may have already occurred (the lie was uttered on the witness stand), but the contextualizing event (the trial itself) is not yet over; the case of "future fraud" deals with events to come, such as the client insisting the day before he is to testify that he will lie under oath. Although there may be debate on what constitutes the "contextualizing events," and thus whether the fraud was past or present, these general categories should be clear enough for purposes of the following discussion.

The first document to be discussed will be ABA Formal Opinion 287.[120] Operating under the old Canons, Opinion 287 dealt head-on with the "lying client" problem. That opinion considered two distinct factual situations. The first involved a client who told his attorney that, during a previous representation, the client had given false testimony on deposition upon which a decree of divorce was later entered. The client had testified that the date of desertion was prior to the required eighteen-month period. The divorce action was, in fact, premature under local law. The client now comes to the same attorney and tells him that his former wife threatens to disclose the true facts unless support money is forthcoming. Clearly this is a case of "past fraud."

The second situation involved a client with a previous criminal

118. M. Freedman, *supra* note 1, at 29.
119. ABA Standards, *supra* note 99, Supplement, at 18.
120. See note 48 *supra*.

record who was told by the sentencing judge: "You have no criminal record, so I will put you on probation." This situation had some variants:

(a) The case where the judge asked the defendant if he had a criminal record and the defendant replied that he did not;
(b) The case where the judge asked the defendant's lawyer whether the defendant had a record.

The second situation, together with its variants, involves questions of "present fraud" and "contextualizing event" being the unfinished sentencing hearing. In both the first and second situations, the question was what should the lawyer do? How should he or she interpret a lawyer's duty to the client and to the court under these circumstances?

The author of the majority opinion, Henry Drinker, considered several of the old Canons. These were:

(1) Canon 37 which stated flatly: "It is the duty of a lawyer to preserve his client's confidence."
(2) Canon 41 which provided: "When a lawyer discovers that some fraud or deception has been practiced, which has unjustly imposed upon the court or a party, he should endeavor to rectify it; at first by advising his client, and if his client refuses to forego the advantage thus unjustly gained, he should promptly inform the injured person or his counsel, so that they may take appropriate steps."
(3) Canon 29 which provided: "The counsel upon the trial of a cause in which perjury has been committed owe it to the profession and to the public to bring the matter to the knowledge of the prosecuting authorities."
(4) Canon 22 which provided: "The conduct of the lawyer before the Court and with other lawyers should be characterized by candor and fairness."

With respect to the first situation, the divorce case, the majority opinion stated that Canon 41 was not relevant, as it was directed to cases where the client had secured an "improper advantage" over another. Neither the state nor the court could be considered an "injured person" within the meaning of Canon 41. This construction ignores, however, the crucial phrase "unjustly imposed on the *court* or a party." Nevertheless, the opinion did conclude that there was a genuine conflict between Canon 29 and Canon 37. Drinker resolved the conflict in favor of Canon 37 because Canon 29 did not specifically require the lawyer to advise the court of the client's perjury, and the question as framed dealt only with the duty to

court and to client, not with duties to the "system" or to "self" or to the *prosecuting authorities.* Weighing heavily the argument that to force a disclosure would seriously undermine the usefulness in having lawyers available to discuss matters with clients in absolute secrecy, the opinion determined that the lawyer should try to persuade the client to disclose the truth willingly. If the client refused to do so, the lawyer should simply have nothing further to do with the client but should tell neither "court" nor "authorities." No construction was given to Canon 29's otherwise plain directive to go to the "authorities" in cases of the discovery of past fraud.

In the second situation, the sentencing case, Drinker's opinion indicates that the lawyer should not volunteer information about the client's lie to the court regarding his past record if the client revealed the information about his record while seeking the lawyer's professional advice. In variant (a), absent such a "privileged communication," the lawyer should make certain the court is not misled into believing that the lawyer is corroborating the client's statement. If asked directly by the court in variant (b), the lawyer should not lie to the court but should retire from the case. Presumably the lawyer should act similarly (not lie, but withdraw) even in the "privileged communication" case, although the opinion deftly ignores that crucial problem.

In a concurring and dissenting opinion, William Jones did not distinguish the "privilege cases" from the non-privileged, maintaining that the obligation to preserve confidences was broader than even the majority suggested. A dissenting opinion, by Wilbur Brucker and William White, claimed that Canons 41, 29, 22, and 15[121] were very clear and should override Canon 37 in both situations.[122]

The majority muddled the handling of the Canons. They are read closely like a statute when useful[123] but their language is ignored when it would be awkward to deal with it.[124] Moreover, the policies underlying the Canons are not specified except in conclu-

121. Old Canon 15 stated: "The office of Attorney does not permit, much less does it demand of him, for any client, violation of law or any manner of fraud or chicane." Canons 29, 41 and 22 are set forth in the text.

122. Although the dissenters rested their position in both situations on a "plain meaning" approach to Canons 29 and 41, they also argued in the second situation that a criminal record was analogous to an essential decision not cited by an opponent, and therefore was something the lawyer must disclose to the court. *See* Formal Opinion No. 280, as published in AMERICAN BAR ASSOCIATION, OPINIONS OF THE COMMITTEE ON PROFESSIONAL ETHICS 618 (1967). The Code, DR 7-106(B)(1), continues and codifies this same obligation.

123. See the construction of "injured person" in Canon 41, discussed in the text.

124. See the construction of "authorities" in Canon 29, discussed in the text.

sory fashion ("confidentiality is supreme") or in double-talk.[125] What the majority accepted as permissible behavior in variant (a) of the sentencing case, where a lawyer stood by silently while a client lied to a judge, is absolutely prohibited by the system and, in my opinion, was absolutely prohibited under Canons 41, 29, 22, and 15. My reasons are rooted in the two fundamental propositions of the adversary system advanced earlier[126] as well as in the well known theory of statutory construction, which applies to all rule interpretations: The specific governs over the general.[127] Moreover, lawyers have been appropriately disbarred or otherwise disciplined for maintaining such silence during an adversary proceeding and failing to do something to rectify the damage done to people and to the system itself by the poison of a lie.[128]

It may be presumptuous to speculate how the majority of the Opinion 287 committee would have handled the paradigm case. But if the lawyer had no duty to disclose the client's lie about his past criminal record in variant (a) of situation 2, then it seems doubtful that he would have been found to have a duty to disclose his client's fraud on the witness stand. Would that mean he must idly sit back and allow his client to lie? Would it also mean, if he received his information in a confidential communication, that he must not interfere if witnesses lie? Is timing a factor here? If he knows beforehand of the false testimony to be given, must he do something? Withdraw perhaps? The sentencing case discussed in Opinion 287 was a case in which the lawyer was presumably surprised by the client's statement. Alas, the Opinion does not attempt to distinguish timing factors; nor does it do more with the problem of competing policies than to acknowledge them. The choice is made on the basis of abstract rule rather than policy. The lawyer may never disclose confidential information because of the damage it may do

125. Compare Drinker's language in the opinion: "We yield to none in our insistence on the lawyer's loyalty to the court of which he is an officer;" and "no client may demand or expect of his lawyer, in the furtherance of his cause, disloyalty to the law whose minister he is (Canon 32) or 'any manner of fraud or chicane,'" with the results in the cases discussed in Opinion 287.

126. See notes 22-50 *supra* and accompanying text.

127. The formulation of this maxim may vary, but the idea is fairly commonplace. *See* Llewellyn, *Remarks on the Theory of Appellate Decision and the Rules or Canons About How Statutes Are to be Construed,* 3 VAND. L. REV. 395, 405 n.44 (1950). In the context of Opinion 287, Canons 29 and 41 should govern over Canon 37, the confidentiality canon.

128. *E.g.,* Committee on Professional Ethics and Conduct v. Crary, 245 N.W.2d 298 (Iowa 1976); *In re* Carroll, 244 S.W.2d 474 (Ky. 1951); *In re* Hargis, 190 S.W.2d 333 (Ky. 1945); *In re* Hardenbrook, 135 App. Div. 634, 121 N.Y.S. 250 *aff'd mem.*, 199 N.Y. 539, 92 N.E. 1086, *appeal denied,* 144 App. Div. 928, 129 N.Y.S. 1126 (1910); *In re* King, 7 Utah 2d 258, 322 P.2d 1095 (1958).

to the attorney-client relationship, but may disclose if the information is not strictly "privileged," no matter the damage to the relationship. Moreover, the lawyer may never lie to the court, no matter what damage is done, but may permit the client to lie to preserve some confidences.

When the American Bar Association promulgated a new Code of Professional Responsibility, effective January 1, 1970, it contained the following provision as DR 7-102(B)(1):

> A lawyer who receives information clearly establishing that:
> His client has, in the course of the representation, perpetrated a fraud upon a person or tribunal shall promptly call upon his client to rectify the same, and if his client refuses or is unable to do so, he shall reveal the fraud to the affected person or tribunal.

In 1974 DR 7-102(B)(1) was amended, adding the following language after the word "tribunal" at the end of the provision: "except when the information is protected as privileged communication." Putting the amendment to one side for the moment, the original DR 7-102(B)(1) would seem clearly to have changed the result in both situations discussed in Opinion 287. Professor Andrew Kaufman, however, considers the amendment in light of Opinion 287 and the Code and rejects this notion:

> ABA Opinion No. 287 resolved, in the particular fact situations presented, the conflict between Canons 6 and 37, and confidentiality canons, on the one hand, and Canons 29 and 41, the perjury and fraud canons, on the other. DR 4-101 and DR 7-102(A)(4) and (B)(1) of the Code of Professional Responsibility replicate the conflict in prohibitions that existed under the Canons. DR 4-101(C)(2), however, allows a lawyer to reveal confidences or secrets whenever "permitted" under the Disciplinary Rules. As originally adopted, DR 7-102(B)(1) did not contain the "except" clause (added in the 1974 amendments) of the current version, and it can be argued that the purpose of the original wording of the provision was to resolve the conflict between confidentiality and disclosure, and hence to change the result in Opinion No. 287, by making DR 7-102(B)(1) and not DR 4-101(B) controlling whenever the former applied.
>
> The Code does not anywhere say that, although among the footnotes to DR 7-102(B)(1) is an ambiguous "but see ABA Opinion 287." Except for a brief reference in footnote 44 to Canon 7, that is the only reference in the footnotes to Opinion 287 although ABA Opinions 155 and 156, which espouse a philosophy contrary to that of Opinion 287, are quoted and cited in several places. On the other hand, *given the unclear wording of Canon 4 and lack of any stated purpose to overturn Opinion 287, it seems more likely that what the drafters of the Code did was to set up essentially the same conflict of principle that*

appeared in the former Canons, thus leaving lawyers, ethics committees, and courts the problem of resolving the conflicts in different factual situations.[129]

Professor Kaufman's conclusion is unsound for several reasons. Kaufman says that the footnote reference to Opinion 287 is "ambiguous." Perhaps, but the introductory signal to the footnote suggests that Opinion 287 was cited as authority which strongly suggests a contrary position.[130] But the important point, of course, is that the footnotes are not to be read as declarative of the views of the draftsmen.[131] For interpretive purposes they are, therefore, irrelevant. Moreover, other evidence supports the conclusion that DR 7-102(B)(1) was originally intended to mean what it plainly says. Freedman himself understood the provision that way and acted swiftly to make certain that it would not be adopted in the District of Columbia, where he was teaching and practicing law.[132] In addition, Formal Opinion 341[133] strongly supports the interpretation of the original language of DR 7-102(B)(1) as being designed to overturn the majority in Opinion 287.

Professor Kaufman's argument against giving the original DR 7-102(B)(1) its plain meaning is that the provision conflicts with Canon 4, thus setting up the same "policy" conflict as the Drinker committee believed existed between the confidentiality canons and the perjury and fraud canons of the previous ABA Canons. But DR 4-101(B)(1) states that a lawyer shall not knowingly reveal a confidence or secret, except "when permitted under DR 4-101(C)." DR 4-101(C)(2) states that a lawyer may reveal confidences or secrets "when permitted under Disciplinary Rules." Thus, when a Disciplinary Rule such as DR 7-102(B)(1) permits the revelation of a confidence, the restrictions of DR 4-101(B)(1) are broken. In other words, in such a situation, there is no obligation under Canon 4 to maintain the confidence. Without the restriction of Canon 4, it can be argued that the original DR 7-102(B)(1) does *mandate* disclosure of past fraud.

It becomes necessary to address this question of the Code's original meaning, not only because the original provision is still in

---

129. A. KAUFMAN, *supra* note 49, at 146-47 (emphasis added).

130. HARVARD LAW REVIEW ASSOCIATION, A UNIFORM SYSTEM OF CITATION 7 (12th ed. 1976).

131. ABA CODE, *supra* note 35, Preamble and Preliminary Statement, n.1.

132. M. FREEDMAN, *supra* note 1, at 28-29. Freedman notes that "the American Bar Association itself recognized the impropriety of requiring a breach of confidentiality," by adding the 1974 amendment requiring no revelation of client fraud "when the information is protected as a privileged communication." *Id.* at 29.

133. See note 55 *supra.*

the Code of so many states,[134] but also because it raises starkly the question of the methods to be adopted in reading the Code. Although not arguing for an unsophisticated "plain meaning" philosophy in interpreting the Code, I would argue that the Disciplinary Rules were meant to be read and interpreted as a statute. The idea behind the Code's tripartite division was to have the Disciplinary Rules serve as the mandatory rules "below which no lawyer can fall without being subject to disciplinary action."[135] Lawyers are supposed to read, understand, and be able to act in reliance upon them. If the Disciplinary Rules are therefore taken seriously as an integrated statute, the original language of DR 7-102(B)(1) must be read as overturning Formal Opinion 287.

The 1974 amendments, however, added "except when the information is protected as a privileged communication" to DR 7-102(B)(1). If the situations discussed in Opinion 287 were to arise for determination under the amended Code, the divorce case (a "past fraud") would have to be distinguished from the sentencing case (a "present fraud"). DR 7-102(B)(1) only deals with questions of "past" fraud. DR 7-102(A)(4) and (7) demand that the lawyer not participate or assist the client in present or future situations involving false testimony. It seems more realistic to treat as "present" fraud cases all perjuries which the lawyer knows about before the trial ends. The perjury is really a "continuing" matter until the trier of fact renders a decision. It therefore would make no difference whether the lawyer "knows" of the perjury before, after, or during the testimony, because the significant event has not yet come to a conclusion. To treat these matters as "past" as soon as they occur would be like refusing to alter a fraudulent prospectus after it has been typed but not printed (or even printed but not distributed) simply because the fraud was discovered after the type was set. It makes no sense. Under this approach, the 1974 amendment to DR 7-102(B)(1) is irrelevant in deciding the paradigm case because DR 7-102(A)(4) and (7) should control.

But if "past fraud" is defined to include perjuries committed during a trial, the obligation of the original DR 7-102(B)(1) would still impose upon the lawyer a duty to correct any fraud occuring at a trial, and would clearly apply in the paradigm case. The original language of DR 7-102(B)(1) clearly mandated the disclosure of fraud by the lawyer to the "affected person or tribunal" if the client re-

134. This is extremely important, for lawyers are bound by their state Codes, not by the ABA Code. As of October 1976, only seven states had adopted the 1974 amendment. *See* note 45 *supra*.

135. ABA CODE, *supra* note 35, Preamble and Preliminary Statement.

fuses to "rectify the same." The 1974 amendment, however, created an ambiguity by "excepting" this mandatory language right out of existence. After the amendment, a lawyer "shall reveal the fraud," except if it was acquired through a "privileged communication"—which is the channel through which most of the lawyer's information on the subject would no doubt travel. As if to perversely insure that the original DR 7-102(B)(1) would be completely emasculated, Formal Opinion No. 341 announced that the phrase "privileged communication" did not refer to information protected by the attorney-client privilege, but actually meant information protected under the broad definitions of "confidences" and "secrets" provided in Canon 4. Opinion 341 argued that the lawyer would be in a bind if he had to obey DR 7-102(B)(1) in face of the conflict with this obligation to maintain the lawyer-client privilege. If this were the problem, and if the word "privilege" were used, it would have been much more logical to limit the exception to the evidentiary matter. Opinion 341 makes clear, however, that its object is to reinstate the "essence" of 287, although it refers only to the "divorce" case, not to the "sentencing" case.

However strained Opinion 341 may be, the 1974 amendment still must mean something. In search of that meaning, it is essential that the Disciplinary Rules be read together for internal consistency. The amendment surely does not change DR 7-102(B)(1) from a "shall reveal" to a "shall not reveal" situation; it merely changes it from a "shall reveal" to a "may reveal." This reading makes sense not only in terms of the actual language used ("shall . . . except"), thus giving some vitality to DR 7-102(B)(1), but it is also consistent with all of the other exceptions to confidentiality listed in DR 4-101(C) which are permissive in nature ("may reveal"). Thus DR 7-102(B)(1) now "permits" disclosure under DR 4-101(C)(2).

So the lawyer is left alone to try to muddle through "past fraud" cases. But as argued above, this does not permit the lawyer to do as he likes in cases similar to the Freedman paradigm. One court has seen the entire area as so confused that a lawyer who plainly would have been disciplined, at least prior to Opinion 287, for knowingly "proceeding normally" in a trial where his client committed a fraud, was not required to "withdraw" from the case or to disclose the fraud. In support of its decision,[136] the Oregon Supreme Court cited Opinion 287 and unamended DR 7-102(B). Subsequent to the inception of the case, the Oregon State Bar issued an Opinion (No. 227) which requires that in the future lawyers withdraw in all

136. *In re* A., 276 Or. 225, 554 P.2d 479 (1976).

similar cases. The court approved that opinion, but considered it unfair to apply to the case at bar.[137]

The chaos is compounded by ABA Informal Opinions 1314 and 1318.[138] In Opinion 1314, the Committee determined that in a criminal case where the client insists upon taking the stand and giving perjured testimony, the lawyer must withdraw or report to the tribunal if the client insists on testifying. This conduct is mandated by DR 7-102(A)(4), (6), and (7). If the lawyer does not know of the perjury "in advance," then the lawyer's primary allegiance is to the client to protect confidentiality. The lawyer, however, is obliged to call upon the client to rectify the fraud and, if the client does not, the lawyer "may withdraw" under DR 2-110(C).[139] Since the Opinion does not discuss "policies," and since the only other citation is to amended DR 7-102(B)(1), it seems the Committee read the "past fraud" provision to apply immediately after the perjury was committed.

In Opinion 1318, the Committee stated that a lawyer should withdraw if a client charged with murder proposes to commit perjury. The Opinion also dealt with a prior perjury committed by the client in another jurisdiction. Since the prior perjury occurred when the lawyer was not representing the client, the attorney could not betray the client's confidence in seeking its rectification. Again the Committee cited DR 7-102(B)(1). Of course, Opinion 1314 would indicate that the "except" clause of the 1974 amendment is also a reason for nondisclosure, even if the lawyer had represented the client in the other matter.

The conclusion to be drawn from ABA Formal Opinion 341 and Informal Opinions 1314 and 1318 is that where a lawyer can withdraw from "lying client" cases, he should. He should make no disclosure, unless he is forbidden to withdraw and the perjury has not yet occurred. Presumably, the duty not to disclose would give the judge enough reason to allow the withdrawal. However, a withdrawal will not always be allowed.[140] Presumably, if the lawyer is not permitted to withdraw or if the perjury "surprises" the lawyer, he still has the problem of how to shape his advocacy. The proposed solutions to Freedman's paradigm show an array of possibilities.

---

137. *Id.* at 233, 554 P.2d at 487.

138. ABA COMMITTEE ON ETHICS AND PROFESSIONAL RESPONSIBILITY, INFORMAL OPINIONS Nos. 1314, 1318 (1975).

139. Presumably the reference is to subpart (C)(1)(b), which states that the lawyer "may withdraw" if the client "personally seeks to pursue an illegal course of conduct."

140. *See* United States v. Gonzales, 435 F.2d 1004 (10th Cir. 1970). See note 103 *supra* and accompanying text.

With the principles advanced and the arguments made herein, it may be possible to advance guidelines for advocacy in various "lying client" situations which are most consistent with the demands of the adversary system. This is the final subject to be considered in this article.

V. The Ethical Role of the Advocate

As a basic philosophical proposition, the adversary system is rooted in a timeless belief that dialogue will yield truth. In recognition of the complexities of life and law, and of the unevenness of talents and education, lawyers act as surrogates for the dialoguing that must exist between contending sides. The philosophy is thwarted and the practicalities made unreal when lying is condoned. Lying cannot be tolerated; the duty of the lawyer-surrogate must include a prohibition against his assisting or permitting a lie to be perpetrated within the system. Moreover, although the system recognizes the need for clients to consult freely and openly with their lawyers without fear that the lawyer may later disclose confidential information to the client's detriment, the system has always recognized significant exceptions to the rule against lawyers breaching such confidences. The system has never condoned lying. Confusion reigns because "silence" can easily be considered a "lie" by omission, and therefore as reprehensible as a deliberate positive falsehood. The system, however, has never recognized silence, as such, to be the equivalent of a lie. Since the lawyer is the client's surrogate, the client's right to silence must be respected. For the same reason, however, when the client lies, the surrogate has a responsibility to the court and to the system to ensure that the lie is exposed. Not to reveal the lie is, therefore, tantamount to assisting the client to misuse the system. It cannot be allowed.

But these are hard sayings. They are hard, paradoxically, because it may mean that a lawyer has to stand by silently and allow an injustice to be perpetrated on another or on the system to the benefit of his client (as in the Williston vignette);[141] or it may mean disclosure of a client's lie to the obvious harm of a client (as must be done in the Freedman paradigm),[142] whether "justice" is done or not.[143] Acceptance of these various outcomes is possible only if the

141. See note 27 *supra* and accompanying text.

142. See note 73 *supra* and accompanying text.

143. Justice presumably would be done if the client were truly "guilty" and the disclosure led to a conviction. Justice would presumably not be done in the Freedman paradigm if an innocent man were convicted because the judge or jury disbelieved the client's entire story because one part of it was fabricated.

idea of one's role as an advocate is properly understood and embraced.

Thus, in the Freedman paradigm, if the client insists on taking the stand and lying, the lawyer has no choice but to disclose the fact of the lie to the court. The lawyer may request a withdrawal from the representation, but permission to grant or deny it lies within the sound discretion of the court. In any event, the determination of the client to lie or the "surprise" lie on the witness stand must each be disclosed, whether withdrawal occurs or not.

The solutions to the situations discussed in Formal Opinion 287 are important. In the divorce case, the lawyer should disclose the past fraud if his client refuses to do so. Because of the 1974 amendment, such conduct is unfortunately no longer mandated under the Code, but it is permitted and every lawyer should disclose because the adversary system demands it. However, in the sentencing case, the lawyer has an obligation of silence if the court simply places the client on probation because of the court's own factual error. This is essentially the Williston case all over again. If the client lies to the judge, the lawyer has an obligation to correct the lie. I see no reason why he cannot request his client to correct the lie first (however awkward that may be to do); the falsehood, however, should not be allowed to stand because it is inconsistent with the adversary system. If the lawyer is asked directly, he may refuse to answer because the request is improper. Of course, this will alert the judge to ask the client or double check somehow, but the result is not the issue. A lawyer's responsibility is the issue.

As Polster argues,[144] and as at least one case reveals,[145] the client may either be dissuaded from testifying or from lying if he does testify, if the lawyer forthrightly explains what he will do when confronted with this impermissible behavior. It should be stressed that counsel should have an objective basis for his belief that the testimony is false,[146] and counsel, as advocate, should give his client the benefit of the doubt;[147] nevertheless, counsel may never permit a lie to be considered as part of the record if he "knows" that it is a lie, no matter the awkwardness of the outcome.

This is essentially a bottom-line position. Different contexts may result in different kinds of advocacy once the advocate understands that his paramount obligation is to make certain that he does not permit a lie to remain on the record for the trier of fact to

144. *See* Polster, *supra* note 98.
145. United States v. Gonzales, 435 F.2d 115 (10th Cir. 1970).
146. *See* United States v. Johnson, 555 F.2d 115 (3rd Cir. 1977).
147. *See ABA Report, supra* note 10, at 1160-61.

consider with the other evidence. The cases indicate that the courts are confused as to how to proceed after the perjury has been made known, but are unanimous in holding the lawyer to a standard consistent with the bottom-line position argued in this article.[148] The confusion is really among judges in terms of the exercise of their discretion. There is no doubt that the matter needs clarification; the clarification should come from the Supreme Court, which can delineate the constitutional parameters of a court's discretion, thereby providing some uniformity. If a court permits counsel to continue without "participation" or permits a "withdrawal" and a new trial, it should not be a matter of overriding concern to the ethical lawyer. Of couse, it will matter *greatly* to the client. As Fuller says,[149] however, the lawyer's primary obligation is to procedures and to institutions; he must fulfill that obligation no matter the cost.

VI. Postscript: Treachery, Substantive Justice, and the Adversary System

Although Section V could stand as a conclusion to this article, I am uncomfortable with so "final" a word as "conclusion," and want to add an additional word or two of postscript. The focus of all problems of lawyers' ethics must be the adversary system. Whatever we believe the characteristics of that system to be, people of good will want it to be the best vehicle for obtaining justice under law that the human mind can construct. We know it does not always produce "justice;" indeed, we know that the role we assign to lawyers in the system at time obstructs justice, just as we know the lawyer's obligation of confidentiality at times thwarts truth. However, we have not obliged lawyers to seek substantive justice directly, but only to play a role in the quest to obtain a system that will produce justice more often and more pervasively than any other system. So to ask the lawyer to play his role in order to achieve justice as he sees it or even "knows" it to be is to ask him to abandon the system we have constructed. Abandonment may be the best thing. Certainly altering the system in one particular or another

148. Some endorse the ABA Standards position. *See* State v. Lowery, 111 Ariz. 26, 523 P.2d 54 (1974). Others seem to believe permission to withdraw must be granted. *See* State v. Robinson, 290 N.C. 56, 224 S.E.2d 174 (1976) (however, the question of whether substitute counsel should be allowed is answered differently). *Compare* Wilcox v. Johnson, 555 F.2d 115, 120 (1977), *with* State v. Robinson, 224 S.E. 2d at 180. One judge felt compelled to disqualify himself from sitting further on the case. This approach was disapproved by the circuit court. *See* Thornton v. United States, 357 A.2d 429 (Ct. App. D.C.), *cert. denied,* 429 U.S. 1024 (1976).

149. *See ABA Report, supra* note 10, at 1162.

may produce better overall results, so experiment we must. Nevertheless, as presently constituted, to ask for a lawyer's effort to produce substantive justice is only to ask ourselves how we want him to play his role.

One thing lingers. In discussing the history of the attorney-client privilege, it was noted that although the honor-treachery grounds have been abandoned as the support and justification of the privilege, the lawyer's concern with the treacherous revelation of a client's confidences has never been buried.[150] This is the heart of the matter. As lawyers, as advocates, we represent clients; we identify ourselves with them; we invite their honesty and candor as we pledge our assistance. So even if an individual does not seem to be our "special friend,"[151] our role-playing demands that we treat the client so. Perhaps, however, we do not understand our total role. We are to be "objective" and "independent"[152] because we have many functions. One of those functions is to be an "officer of the court." It is easy to advance reasons why Freedman's position is wrong, even if one were not trying to find out what the system demands. To accept Freedman's position is to unleash the possibility that clever lawyers would be able to "create" stories for or in cooperation with their clients to the detriment of us all. We cannot abide that because we, as members of society, feel betrayed by giving lawyers a monopoly on the practice of law and then watching helplessly as they assist the wicked or the undeserving, not in speaking their truths for us to judge, but in speaking their falsehoods for their own benefit. Of course, as Freedman's paradigm case makes clear, the innocent may suffer too, if no help in lying is afforded. I am willing to take my chances on whose side the balance will tip.

Moreover, we need general rules to live by. I have proposed a start—or rather Freedman did: he proposed we determine lawyers' ethics in light of the demands of the adversary system. I believe I have suggested more significant demands than those advanced by Freedman. If there is the possibility of treachery to the client on the one hand, there is also the possibility of treachery to the system on the other. There is no dilemma unless we look simply to results and become act utilitarians. As lawyers, we are obliged to be rule utilitarians. If the results produced are not to our liking, the system should be changed. However, I see little hope for improvement until we better understand the system itself.

---

150. See text accompanying note 64 *supra*.

151. The phrase is Charles Fried's. *See* Fried, *The Lawyer as Friend: the Moral Foundations of the Lawyer-Client Relation*, 85 YALE L.J. 1060 (1976).

152. ABA CODE, *supra* note 35, Canon 5.

I have proposed a modest start. Lawyers must not allow lies to stand. But what should judges do in response to a disclosure?

Let the dialogue continue.

Utah Law on Municipal Boundary Changes—Anarchy Among Modern City-States

State legislatures have the power to create and abolish municipal corporations[1] within constitutional and statutory limits.[2] As an incident to this power, legislatures may alter municipal boundaries.[3] When a legislature possesses this nondelegable authority,[4] its enactments determine how and under what policies municipal boundary changes[5] are made. To provide a uniform and consistent scheme for accomplishing municipal boundary changes, a state legislature must establish basic policy guidelines. More particularly, a legislature must determine who should be allowed to initiate, supervise, decide, appeal, and veto municipal boundary changes. Statutes allowing only a select class of individuals, such as property owners, to participate in the process, or statutes requiring separate majorities, have been constitutionally challenged in recent years. In addition, many current statutes are subject to criticism for retaining inconsistent and antiquated policies.

The Utah Legislature ranks among the most neglectful in providing adequate statutory guidance for municipal boundary

---

1. *E.g.*, UTAH CONST. art. XI, § 5. *See* 1 C. ANTIEAU, MUNICIPAL CORPORATION LAW §§ 1.01, 2.00 (1975); 2 E. MCQUILLIN, THE LAW OF MUNICIPAL CORPORATIONS § 4.03 (rev. 3d ed. 1966).

2. *See* Hunter v. City of Pittsburgh, 207 U.S. 161, 178-79 (1907). For an example of when the legislature will be constrained by constitutional limits on its power to alter municipal boundaries, see Gomillion v. Lightfoot, 364 U.S. 339 (1960) (state could not, under due process and equal protection clauses, alter municipal boundaries to exclude from a city all but four or five Blacks). In 1869, Chief Justice Dillon of the Supreme Court of Iowa stated:

> Municipal Corporations owe their origin to, and derive their powers and rights wholly from the legislature. It breathes into them the breath of life, without which they cannot exist. As it creates, so it may destroy. If it may destroy, it may abridge and control. Unless there is some constitutional limitation on the right, the legislature might, by a single act, if we can suppose it capable of so great a folly and so great a wrong, sweep from existence all of the municipal corporations in the State, and the *corporation* could not prevent it. We know of no limitation on this right so far as the corporations themselves are concerned. They are, so to phrase it, the mere *tenants at will* of the legislature.

City of Clinton v. Cedar Rapids & Mo. River R.R., 24 Iowa 455, 475 (1869). *See also* J. WINTERS, STATE CONSTITUTIONAL LIMITATIONS ON SOLUTIONS OF METROPOLITAN AREA PROBLEMS 6-10 (1961).

3. *See* Hunter v. City of Pittsburgh, 207 U.S. 161, 178-79 (1907); *In re* Peterson, 92 Utah 212, 216-17, 66 P.2d 1195, 1197-98 (1937); 2 E. MCQUILLIN, *supra*, note 1, at §§ 7.10, 7.24.

4. *See* 1 C. ANTIEAU, *supra* note 1, at § 1.03.

5. In this Note the phrase "municipal boundary changes" will be used to include the functions of incorporation, annexation, disconnection, consolidation, and dissolution unless otherwise indicated. The formation and alteration of special districts, though sometimes governed by similar statutory provisions, will not be included in the usage of the term.

changes.[6] In 1977 the Utah Legislature adopted the first part of a Utah Municipal Code, as an attempt to modernize and codify Utah laws relating to municipalities.[7] The legislation, however, fails in its purpose, remaining devoid of policy guidelines on municipal boundary changes. Moreover, the legislation is constitutionally suspect on several grounds. This Note examines the Utah statute and the constitutional and policy problems it raises. It also recommends ways to avoid constitutional problems relating to citizen participation in municipal boundary changes, to add clarity and consistency to municipal boundary change procedures, and to facilitate sound municipal development through more adequate policy guidelines.

I. UTAH MUNICIPAL CODE

Chapter 2 of the Utah Municipal Code covers the processes of incorporation, annexation, disconnection, consolidation, and dissolution[8] by creating procedures for the initiatiion, supervision, decision, and appeal of such actions.

A. Initiation

The requirements for initiating municipal boundary changes vary with each type of change. To initiate the incorporation of a city, a petition must be signed by not less than one hundred registered voters in the territory to be incorporated;[9] for towns, by a majority of electors.[10] Dissolution, the converse function, requires a petition of twenty-five percent or more of the registered voters.[11] To initiate an annexation, a petition must be signed by "a majority of the owners of real property and the owners of at least one third in

6. Writing on the need for more policy guidelines in Utah's annexation and incorporation statutes, Dean Jefferson B. Fordham, Distinguished Professor of Law, University of Utah, said, "I have examined related legislation of numerous states and I say without hesitation that the Utah statutes are the most superficial and inadequate I have ever seen. Neither sets any positive standards at all." Fordham, *Meager Utah Statutes Hinder Municipal Incorporation*, Salt Lake Tribune, Aug. 5, 1977, § A, at 23, col. 1.

7. UTAH CODE ANN. §§ 10-1-101 to 3-1228 (Supp. 1977). For a brief summary and analysis of the Utah Municipal Code see *Utah Legislative Survey*—1977, 1977 UTAH L. REV. 521, 522.

The remaining two parts of the Utah Municipal Code will be introduced in the regular sessions of the 43rd and the 44th Utah Legislature. Interview with Michael T. McCoy, Legal Counsel for Utah League of Cities and Towns, in Salt Lake City, Utah (Oct. 8, 1977). *See also* letter from Michael T. McCoy to Municipal Officials of Utah (Mar. 29, 1977).

8. UTAH CODE ANN. §§ 10-2-101 to 113 (incorporation), -401 to 404 (annexation), -501 to 509 (disconnection), -601 to 614 (consolidation), -701 to 712 (dissolution) (Supp. 1977).

9. *Id.* § 10-2-101.

10. *Id.* § 10-2-109. *See also* note 26 *infra*.

11. *Id.* § 10-2-701.

value of the real property" in the territory to be annexed.[12] Noticeably, the statute does not require the property owners to be registered voters or even inhabitants of the territory to be annexed. A petition to disconnect, however, must be signed by a "majority of the real property owners in any territory within and lying on the borders of any incorporated municipality,"[13] and the district court must find that "the petition was signed by a majority of the registered voters of the territory concerned."[14] Consolidation may be initiated by petition "signed by at least ten percent of the registered voters in each of the municipalities to be included," or by "[r]esolutions passed by the governing bodies of the municipalities."[15]

B. Supervision

No central agency exists in Utah to coordinate the various supervising bodies and to review the overall effect of boundary change proposals. The board of county commissioners receives and verifies the petitions for incorporation[16] and consolidation.[17] Petitions for annexation are filed with the municipal governing body[18] while those for dissolution[19] and disconnection[20] are filed with the district court of the county in which the territory is located. These various supervisory bodies, when applicable, call public hearings,[21] set times and places for, and publish notice of, elections,[22] canvass ballots,[23] and may order levy of taxes to pay past indebtedness.[24]

---

12. *Id.* § 10-2-401. For a slightly dated summary of the annexation laws in all fifty states, as well as the "Basic Principles for a Good Annexation Law" as developed by the National League of Cities, see DEPARTMENT OF URBAN STUDIES, NATIONAL LEAGUE OF CITIES, ADJUSTING MUNICIPAL BOUNDARIES (1966).

13. UTAH CODE ANN. § 10-2-501 (Supp. 1977).

14. *Id.* § 10-2-502. Reading sections 10-2-501 and 10-2-502 together seems to indicate that persons who sign petitions to disconnect must be both "real property owners" and "registered voters." Though in many cases a person would meet both requirements, corporate property owners would be excluded from petitioning. One wonders if the legislature actually intended a dual-eligibility requirement and, if so, why it was adopted for only this boundary change process.

15. *Id.* § 10-2-601.

16. *Id.* § 10-2-101. *See also* note 26 *infra*.

17. *Id.* § 10-2-601.

18. *Id.* § 10-2-401. For a comparative analysis of various methods of supervising and reviewing annexation proposals, see Woodroof, *Systems and Standards of Municipal Annexation Review: A Comparative Analysis*, 58 GEO. L.J. 743 (1970).

19. UTAH CODE ANN. § 10-2-701 (Supp. 1977).

20. *Id.* § 10-2-501.

21. *Id.* § 10-2-606 (consolidation).

22. *Id.* §§ 10-2-102 (incorporation), -702, -703 (dissolution).

23. *Id.* § 10-2-104 (incorporation).

24. *Id.* §§ 10-2-506 (disconnection), -706 (dissolution).

C. *Decision*

Approval by a majority of voters in an election effects the incorporation of a city and the consolidation, or dissolution, of a city or town.[25] The incorporation of a town, since the petition was signed by a majority of the electors, needs only the approval of the county commission.[26] A two-thirds vote of the municipal governing body approves an annexation;[27] the district court determines a disconnection.[28]

D. *Appeal*

The Municipal Code includes a provision, by negative implication, to "challenge" an incorporation or to "protest" an annexation. If "a municipality shall have levied and collected a property tax for two or more years following its incorporation and no person has *challenged* the existence of the municipality in the district court . . . the municipality shall be conclusively presumed to be a lawfully existing and incorporated municipality."[29] Annexations are deemed conclusive if the inhabitants pay the property tax levied and "no inhabitants of the territory *protests* [sic] the annexation during the year following the annexation."[30] Noticeably, the provision relating to protest of an annexation is devoid of instruction on how, where, or upon what grounds such protest may be brought.[31]

---

25. *Id.* § 10-2-106 (incorporation), -610 (consolidation), -705 (dissolution).

26. *Id.* § 10-2-109. In Cottonwood City Electors v. Salt Lake County Bd. of Comm'rs, 28 Utah 2d 121, 499 P.2d 270 (1972), the Utah Supreme Court held that the Board had complete discretion to approve or disapprove an incorporation petition unless the decision was "so wholly without reason as to be capricious and arbitrary." *Id.* at 125, 499 P.2d at 273. For an analysis of that case and of some problems resulting from Utah's municipal incorporation statute similar to those raised in this Note, see Comment, *Utah's Future Municipalities—Incorporation by Chance or Design?*, 1972 UTAH L. REV. 597.

27. UTAH CODE ANN. § 10-2-401 (Supp. 1977).

28. *Id.* § 10-2-505.

29. *Id.* § 10-2-113 (emphasis added). For background on appeals to incorporation, see Mandelker, *Municipal Incorporation on the Urban Fringe: Procedures for Determination and Review*, 18 LA. L. REV. 628, 652-58 (1958).

30. UTAH CODE ANN. § 10-2-403 (Supp. 1977) (emphasis added).

31. Meager common law on the problem does not offer much more guidance. In Bradshaw v. Beaver City, 27 Utah 2d 135, 493 P.2d 643 (1972), several "taxpayers and residents within Beaver City" sought to enjoin the annexation of a tract of land on the contention that the annexation was "arbitrary, unreasonable, capricious and not done in accordance with law and the prerogatives of the defendant City Council." *Id.* at 136, 493 P.2d at 644. In affirming a summary judgment for the city, the Utah Supreme Court stated:

> The courts are and should be reluctant to intrude into the prerogative of the legislative branch of government, and will interfere with such action only if it plainly appears that it is so lacking in propriety and reason that it must be deemed capricious and arbitrary, or is in excess of the authority of the legislative body.

Id. at 137, 493 P.2d at 645.

Incorporations may be challenged by any "person," which by definition includes public and private entities,[32] while annexation protests are limited to "inhabitants."[33]

II. Citizen Participation in Municipal Boundary Changes

In the exercise of their power to determine how municipal boundary changes are made, state legislatures must decide how community interests in boundary changes can be voiced and who should represent those interests. Effective expression of citizen concerns must be provided because a democracy values citizen participation and because boundary changes primarily affect people—those who live and work in the municipality, enjoy its services, and support its programs through taxation.

Some states have met the need for citizen participation by granting citizens the right to initiate boundary changes,[34] to present arguments for or against proposals in public hearings,[35] and to make final decisions by vote.[36] In addition, some states have apparently considered property owners to be most directly affected by boundary changes and have granted them special privileges in initiating changes and making final determinations.[37] Recently, such special distinctions have given rise to constitutional challenges on equal protection grounds.

A. Property Ownership Prerequisites to Participation: Constitutional Concerns

In a line of cases,[38] beginning with *Kramer v. Union Free School District No. 15,*[39] the United States Supreme Court has invalidated,

---

32. Utah Code Ann. § 68-3-12 (1968).

33. *Id.* § 10-2-403 (Supp. 1977).

34. *See, e.g.,* Alaska Stat. §§ 29.18.070, .68.040, .68.510 (1972); Cal. Gov't Code §§ 34303, 35000, 35551, 35704 (West 1968 & Supp. 1977); Utah Code Ann. §§ 10-2-101, -109, -401, -501, -601, -701 (Supp. 1977); Va. Code §§ 15.1-946, -966, -1034, -1073 (1973 & Supp. 1977).

35. *See, e.g.,* Alaska Stat. §§ 29.18.080, .68.070, .68.550 (1972); Cal. Gov't Code §§ 34310, 35007, 35555 (West Supp. 1977); Utah Code Ann. §§ 10-2-501 to 505, -606 (Supp. 1977); Va. Code §§ 15.1-1036, -1041 (1973).

36. *See, e.g.,* Alaska Stat. §§ 29.18.110, .68.090, .68.570 (1972); Cal. Gov't Code §§ 34322, 35116, 35556, 35704 (West 1968); Utah Code Ann. §§ 10-2-102, -607, -702 (Supp. 1977); Va. Code §§ 15.1-948, -969, -1075 to 1077 (1973 & Supp. 1977).

37. *See, e.g.,* Mich. Stat. Ann. § 5.2085 (Supp. 1977); Utah Code Ann. §§ 10-2-401, -501 (Supp. 1977); Wis. Stat. Ann. § 66.014(2) (West Supp. 1977).

38. *See* Hagman & Disco, *One-Man One-Vote as a Constitutional Imperative for Needed Reform of Incorporation and Boundary Change Laws,* 2 Urban Law. 459, 461-64 (1970).

39. 395 U.S. 621 (1969). For an analysis of *Kramer,* see Comment, *Limitations on the*

on equal protection grounds, state statutes that allow only property owners to participate in certain types of elections. In *Kramer,* the Court applied a two-tier test to an equal protection challenge of a statute that required voters in school district elections to be either (1) owners or lessees of property within the district, or (2) the parent of a child enrolled in a local district school.[40] Under the test, if a provision of law creates a "suspect classification," such as race,[41] or affects "fundamental rights," such as the right to vote,[42] the provision is strictly scrutinized and the state is required to show a compelling interest to sustain the provision.[43] If a provision does not create a "suspect classification" and does not affect "fundamental rights," the state must only show that it reasonably accomplishes some state purpose.[44] The *Kramer* Court, finding that voting rights are fundamental, held the strict scrutiny test appropriate. The Court found the statute inadequate because it did not sufficiently demarcate between those interested in school district elections and those not interested. According to the Court in *Kramer,* a statute limiting the vote to one class of voters would satisfy the strict scrutiny test only where "all those excluded are in fact substantially less interested or affected than those the statute includes."[45]

The California Supreme Court, in *Curtis v. Board of Supervisors,*[46] went further than *Kramer* and struck down a statute that granted property owners of a majority of the assessed property

---

Voting Franchise and the Standard of Kramer v. Union Free School District No. 15, 1970 UTAH L. REV. 143.

40. *See* 395 U.S. at 623.

41. Korematsu v. United States, 323 U.S. 214 (1944).

42. Dunn v. Blumenstein, 405 U.S. 330 (1972).

43. *See* G. GUNTHER, CASES AND MATERIALS ON CONSTITUTIONAL LAW 658-59 (9th ed. 1975).

44. *Id.* Recent cases have criticized the two-tier test because too often the decision whether to apply the strict scrutiny test decides the ultimate question as to whether the provision violates the equal protection clause. A "newer" equal protection analysis beginning to be applied in Supreme Court decisions may limit the "fundamental areas" in which "strict scrutiny" is applied and at the same time increase judicial inquiry into other areas by strengthening the "minimal rationality" standard. Under this analysis to sustain a statute, the state must show "that legislative means . . . substantially further legislative ends." Gunther, *The Supreme Court 1971 Term—Foreward: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection,* 86 HARV. L. REV. 1 (1972). In addition, the analysis asks the Court "to assess the rationality of the means in terms of the *state's* purposes, rather than hypothesizing conceivable justifications on its own initiative." G. GUNTHER, *supra* note 43, at 686-87. *See also* note 70 *infra* and accompanying text.

45. 395 U.S. at 632.

46. 7 Cal. 3d 942, 501 P.2d 537, 104 Cal. Rptr. 297 (1972). *See also* Levinsohn v. City of San Rafael, 40 Cal. App. 3d 656, 115 Cal. Rptr. 309 (1974) (invalidating property owner protest of annexation).

value the power to *prevent* an election on incorporation by filing an appropriate protest. Although the right to vote was not denied outright, the court applied the strict scrutiny standards of *Kramer* because the "veto power" touched upon and burdened the right to vote.[47]

The Delaware Chancery Court, in *Kelley v. Mayor of Dover*,[48] also applied the strict scrutiny test where the right to vote was only indirectly affected. The court invalidated a statute that weighted votes according to the assessed value of land in connection with annexation elections because the provision diluted the vote of those who owned little in land value. Though not bound by the California decision, the Delaware court found guidance in the basic principles *Curtis* enunciated:

> The philosophic reach of the decisions of the United States Supreme Court has been to afford to each individual citizen a maximum democratic participation in political matters upon an equal basis. . . .
>
> . . . To frustrate the endeavor of individuals to fix the unit of their local governance and to repose that power in land, not people, would be to stifle that self-determination.[49]

This principle of "maximum democratic participation in political matters upon an equal basis" argues most strongly against the acceptability of statutes that limit participation in municipal boundary changes to owners of real property.

The doctrine of *Kramer* and its progeny[50] should be considered carefully by states, such as Utah, where concerns for equal protection are not entirely obviated.[51] The Utah statute allows only owners

---

47. Curtis v. Board of Supervisors, 7 Cal. 3d 942, 953, 501 P.2d 537, 544, 104 Cal. Rptr. 297, 304 (1972).

48. 314 A.2d 208 (Del. Ch. 1973), *aff'd* 327 A.2d 748 (Del. 1974).

49. Curtis v. Board of Supervisors, 7 Cal. 3d 942, 965-66, 501 P.2d 537, 553-54, 104 Cal. Rptr. 297, 313-14 (1972), *quoted in* Kelley v. Mayor of Dover, 314 A.2d 208, 213-14 (Del. Ch. 1973).

50. *See* Phoenix v. Kolodziejski, 399 U.S. 204 (1970) (general obligation bond); Cipriano v. City of Houma, 395 U.S. 701 (1969) (invalidating statute that allowed only property owners to vote on issuance of revenue bond); Cypert v. Washington County School Dist., 24 Utah 2d 419, 473 P.2d 887 (1970). In *Cypert* the Utah Supreme Court reluctantly adopted the holding of the *Phoenix* case.

> Notwithstanding our emphatic disagreement with the majority in the Phoenix case, we realize that it is for the present to be recognized as the law; and that as such it renders those aspects of the state constitution and statutes inoperable insofar as they require that only property taxpayers be permitted to vote in such bond elections.

Id. at 422, 473 P.2d at 890. *But see*, Salyer Land Co. v. Tulare Water Storage Dist., 410 U.S. 719 (1973) (sustaining statute that limited voting in special water storage district election to property owners).

51. Equal protection problems similar to those involved in municipal boundary changes

of real property to initiate annexation[52] and disconnection[53] proceedings. In addition, the statute weights the significance of each petition signature by requiring signatures from owners of one-third the value of the land.[54] A broad reading of *Kramer* may include as fundamental interests the right to "participate in political affairs,"[55] and not just the right to vote. This interpretation may cause the courts to apply the strict scrutiny test to the Utah provisions, even though no vote is involved. If the strict scrutiny test were applied, it is very doubtful that the state could show a compelling interest sufficient to sustain the requirement of property ownership.[56]

B. Property Ownership Prerequisites to Participation: Policy Concerns

Even if the statute withstood equal protection challenges, however, it may be wanting for policy reasons. The requirement of prop-

---

exist in other statutory provisions. Utah, for example, allows owners of forty percent of the value of real property within a proposed service area to prevent further consideration of the creation of the service area. UTAH CODE ANN. § 17-29-9 (Supp. 1977). A Utah Supreme Court decision in May 1976 invalidated a Salt Lake County scheme that funded services for unincorporated areas of the county from general revenue funds. Salt Lake City Corp. v. Salt Lake County, 550 P.2d 1291 (Utah 1976). Following the decision, the county attempted to establish a service area to provide funds for the unincorporated areas. The proposal was blocked, however, by Kennecott Copper Corporation in conjunction with several other large corporations. At the time, Kennecott Copper's assessed property value equaled more than thirty percent of the assessed value of land in the unincorporated areas of the county (based on the 1975 valuation). Interview with Gerald H. Kinghorn, Assistant Salt Lake County Attorney (Jan. 10, 1978).

52. UTAH CODE ANN. § 10-2-401 (Supp. 1977).

53. *Id.* § 10-2-501.

54. *Id.* § 10-2-401.

The requirement that owners of one-third of the assessed value of the land must sign the annexation petition puts much control in the hands of corporate and business land owners. No exact figures are available to show a comparison in assessed valuation of business to residential property, but a rough figure can be obtained by comparing state property assessments, which cover mines, utilities, and public transportation companies, with county property assessments, which cover all other property situated in the county. In 1976, for example, state assessments comprised twenty-five percent of the total assessed valuation in Salt Lake County. This figure underestimates business property, however, because county assessments include shopping malls, retail outlets, and other businesses. Figures prepared by Nelson Williams, Salt Lake County Auditor's Office.

55. Kramer v. Union Free School Dist. No. 15, 395 U.S. 621, 626 (1969).

56. The Utah provision could also face serious challenge under "newer" equal protection analysis, without strict scrutiny, because the legislature has failed to articulate a "state purpose" to support the property ownership requirement. See note 44 *supra*.

In *Curtis*, the California Court upheld provisions allowing incorporation proceedings to be commenced by petition of owners of "25 percent of the assessed value of land." But it specifically noted "that serious constitutional issues may arise from the failure of section 34303 to permit incorporation proceedings to be initiated by 25 percent of the residents, without regard to their ownership or value of land." Curtis v. Board of Supervisors, 7 Cal. 3d 942, 964 n.30, 501 P.2d 537, 553 n.30, 104 Cal. Rptr. 297, 313 n.30, (1972).

erty ownership to participate in decision-making processes involving municipal boundary changes is seemingly motivated by property tax concerns. The validity of such concerns is questionable. The courts have noted that persons other than land owners are affected by property taxes.[57] Tax increases will be passed along to the tenant in the form of rent increases. The costs of goods and services also reflect increases in property taxes.

An estimate of the size of one group of residents who may be affected by property taxes and yet excluded from those functions requiring property ownership can be extrapolated from statistics gathered from the Salt Lake City metropolitan area during the 1970 census. The census shows that 35.7% of the households in the Salt Lake area consist of renter-occupied housing units.[58] Assuming an equal distribution of voters between renter-occupied and owner-occupied housing units, approximately 110,000 registered voters would be incapable of signing petitions for annexation or disconnection.[59] This exclusion may be even more disturbing given the fact that 60% of households with an annual income below the poverty level occupy rented housing units.[60]

More importantly, a policy favoring property taxpayers is of doubtful validity when property taxes no longer provide the exclusive revenue sources of municipalities. State and federal revenue sharing, grants, and appropriations actually exceed contributions from property taxes in city financing. In Salt Lake County, as an illustration, the general property tax in 1976 constituted less than a third of the revenues in the general fund.[61] Other sources of funds included principally sales taxes, service charges, and federal and state grants. From 1972 to 1976, sales taxes jumped from 11% to 20% of general fund revenues. During the same period, property taxes dropped from 58% to 31% of total general fund revenues. The increased reliance on sales and other taxes may indicate a growing trend to rely less on real property taxes than on taxes levied on all residents and other users of municipal services.

---

57. *See, e.g.*, Phoenix v. Kolodziejski, 399 U.S. 204, 210-12 (1970); Kramer v. Union Free School Dist. No. 15, 395 U.S. 621, 630 (1969); Curtis v. Board of Supervisors, 7 Cal. 3d 942, 960-61, 501 P.2d 537, 550, 104 Cal. Rptr. 297, 310 (1972).

58. BUREAU OF THE CENSUS, U.S. DEP'T. OF COMMERCE, 1970 CENSUS OF POPULATION: GENERAL SOCIAL AND ECONOMIC CHARACTERISTICS: UTAH 46-174 (1972).

59. Based on the number of registered voters for the 1976 general election, as recorded by the Salt Lake County election clerk and compiled by the authors.

60. BUREAU OF THE CENSUS, *supra* note 58.

61. All figures in this paragraph are from data compiled by Nelson Williams of the Salt Lake County Auditor's Office from the Annual Financial Reports of Salt Lake County from 1972 to 1976 prepared by Brunson, Pickett, Osborne & Co., CPA and Peat, Marwick, Mitchell & Co., CPA. *See* Salt Lake County General Fund Revenues, at Appendix *infra*.

Since statutes that stress property ownership may no longer reflect modern municipal realities, the state legislature should reevaluate the emphasis placed on ownership of land. In considering boundary changes, land owners should be given no greater preference than is given other interested citizens.

C. *Participation of Nonresident Property Owners*

The existence of nonresident and corporate owners of property further complicates the policy of basing participation in boundary decisions upon land ownership.[62] Courts, however, have upheld statutes that deny nonresident property owners the right to vote on boundary changes. In *Erven v. Riverside County Board of Supervisors,*[63] a California court held that a statute denying a vote to nonresident property owners did not violate the equal protection clause. Indeed, the court declared that "absent demonstration of a compelling state interest for so requiring, extension of voting rights to non-resident landowners equivalent to that accorded residents would result in an unconstitutional dilution of the voting rights of residents."[64] Nonresident property interests do, however, retain some protection if state legislatures require public hearings open to all interested parties.

D. *Balancing Local Autonomy and Area-Wide Interests*

The state legislature faces a difficult problem when it grants final resolution of consolidation and annexation proposals to voters. How is the legislature to balance the interests of towns and cities in preserving their local autonomy with the area-wide interest in efficiently and effectively providing services to all members of the community at large? If determination of the question is made by an aggregate majority of all voters involved in the consolidation, residents of smaller towns and cities may suffer from loss of community identity, greater taxation, and less influence on the provider of public services.[65] Conversely, if a concurrent majority of each municipality is required for consolidation, residents of particular towns or cities may attempt to preserve benefits for their own communities,

---

62. See note 54 *supra.*

63. 53 Cal. App. 3d 1004, 126 Cal. Rptr. 285 (1975).

64. *Id.* at 1018, 126 Cal. Rptr. at 294.

65. Concurrent majority provisions can be found in many state statutes including Utah. *E.g.,* CAL. GOV'T CODE § 35722 (West 1968); UTAH CODE ANN. § 10-2-610 (Supp. 1977); VA. CODE § 15.1-1103 (Supp. 1977); WASH. REV. CODE ANN. § 35-10-240 (Supp. 1976). *See* Note, *Stumbling Giants—A Path to Progress through Metropolitan Annexation,* 39 NOTRE DAME LAW. 56, 65 (1963).

to the detriment of others, by expropriating high revenue-producing areas and preventing efficient distribution of public services.[66]

Underlying the decision to require approval by each city in a consolidation is the concern for equal protection that arises when voters in one city disapprove the consolidation, while the overall majority approve it. In effect, the voters of the disapproving city are given greater voting strength than the other voters. Because the dilution of voting power affects the right to vote, some courts, under the *Curtis* rationale, have applied the strict scrutiny test to determine whether the state has a compelling interest in separate majority provisions. Although some cases have rejected separate majority requirements for the amendment of state constitutions[67] and reapportionment of state legislative bodies,[68] one federal court upheld the requirement of separate majorities in a consolidation of school districts.[69]

Recently the Supreme Court, in *Town of Lockport v. Citizens for Community Action,*[70] reversed a lower federal court decision invalidating a New York law requiring separate approval of a new county charter by city and non-city voters. After discussing earlier cases dealing with the classification of voters into interested and non-interested groups,[71] Justice Stewart concluded that the proper decision of the case depended upon two questions: "whether there is a genuine difference in the relevant interests of the groups that the state electoral classification has created; and if so, whether any resulting enhancement of minority voting strength nonetheless amounts to invidious discrimination in violation of the Equal Protection Clause."[72] After finding "a genuine difference in the relevant interests" of city and non-city voters,[73] the Court determined that the "differences [were] sufficient under the Equal Protection Clause to justify the classifications."[74] In dicta, the Court stated

66. *Stumbling Giants, supra* note 65, at 57-59.

67. State v. State Canvassing Bd., 78 N.M. 682, 437 P.2d 143 (1968). *See also* Hill v. Stone, 421 U.S. 289 (1975) (invalidating provision for separate approval by all voters and by voters who had rendered property to be taxed); Hagman & Disco, *supra* note 38, at 470-72.

68. Holt v. Richardson, 238 F. Supp. 468 (D. Hawaii 1975). *See also* Curtis v. Board of Supervisors, 7 Cal. 3d 942, 961-62, 501 P.2d 537, 550-51, 104 Cal. Rptr. 297, 310-11 (1972).

69. Keane v. Golka, 304 F. Supp. 331 (D. Neb. 1969).

70. 97 S. Ct. 1047 (1977).

71. *Id.* at 1053, *citing* Hill v. Stone, 421 U.S. 289 (1975); Salyer Land Co. v. Tulare Water Storage Dist., 410 U.S. 719 (1973); Phoenix v. Kolodziejski, 399 U.S. 204 (1970); Kramer v. Union Free School Dist. No. 15, 395 U.S. 621 (1969); Cipriano v. City of Houma, 395 U.S. 701 (1969).

72. 97 S. Ct. at 1053.

73. *Id.* at 1054-55.

74. *Id.* at 1055.

that in annexation proceedings the different interests of residents in the annexing city and residents of the area to be annexed are "readily perceived."[75] By analogy, "in terms of recognizing constituencies with separate and potentially opposing interests," the annexation decision "is similar in impact to the decision to restructure county government in New York."[76] Finally, the Court granted the statute "the presumption of constitutionality to which every duly enacted state and federal law is entitled"[77] and found no violation of equal protection guarantees. The Court at no time mentioned the strict scrutiny standard nor inquired into compelling state interests.[78]

Although separate majorities for consolidation may be constitutionally permissible, pertinent policy considerations argue against such requirements. By allowing suburban voters a veto, the requirements virtually assure the success of defensive incorporation against the expansion of core cities. In many regions of the country, defensive incorporation[79] has been used as a means of preserving tax advantages to residents of small towns surrounding larger core cities. Although the residents of the smaller towns use the core city's services—roads, libraries, museums—for work and leisure, they pay no property taxes to the central city.[80] The central city, in effect, provides services for the entire greater metropolitan area, but receives only limited funds from the residents outside the city limits. The central city's growth into new tax-producing areas in search of funds to provide for badly needed services is stifled because of the statutory veto to consolidation granted the residents of the smaller surrounding towns.

This statutory veto to consolidation also allows localities to take the offensive and acquire by incorporation highly productive tax areas, such as shopping malls, factories, and other businesses, thus retaining the tax revenue benefits for themselves to the exclusion of the surrounding suburbs as well as the central city. In Utah, where municipalities receive significant funds from property taxes

75. *Id.*

76. *Id.*

77. *Id.* at 1056.

78. *See* Kramer v. Union Free School Dist. No. 15, 395 U.S. 621, 634 (1969) (Stewart, J., dissenting). Justice Stewart criticized the majority's application of the "close scrutiny" standard to voting requirements as well as the majority's refusal to grant the requirements the general presumption of constitutionality usually accorded to state statutes. *Id.* at 639. *See also* note 44 *supra*.

79. For a full discussion of defensive incorporation and related issues, see *Stumbling Giants*, *supra* note 65.

80. *Id.* at 57-58.

and sales taxes allocated on a point-of-sale basis,[81] such incorporations pose a significant threat. A side effect of such actions already troubles the Salt Lake County area. As cities are incorporated and annexations occur to acquire tax-producing areas, the county finds itself increasingly left with high-cost service areas and a diminished tax base.[82] The state legislature must address these problems and decide to what extent it can afford to allow local autonomy[83] to rule in consolidation decisions.

III. UNIFORMITY

To implement basic policy decisions effectively, the state legislature must apply them uniformly. Jurisdictional conflicts, inconsistent wording, differing requirements and procedures can only create confusion, prevent successful municipal development, and impede fulfillment of legislative objectives.

The Utah Municipal Code illustrates the problems that arise when several agencies are given jurisdiction over different types of municipal boundary changes. The Code allocates the supervision of procedures relating to incorporation and consolidation to the county commission,[84] those relating to annexation to the governing body of the municipality,[85] and those relating to disconnection and dissolution to the district court.[86] Problems can occur under such a system when two or more bodies are considering different actions relating to the same territory. May a city annex territory that is being considered by the county in an incorporation proposal? The question

---

81. *See* Appendix *infra*. A point-of-sale distribution means that sales taxes are allocated to the local government within whose boundaries the taxes were collected.

82. Interview with Gerald H. Kinghorn, Assistant Salt Lake County Attorney (Jan. 10, 1978). Mr. Kinghorn suggested further that continued reduction in the county's tax base without corresponding reductions in high-cost service areas could eventually result in suburban slums.

83. The freedom of the individual to choose whether or not he wishes to be included within the jurisdiction of a municipal government is proclaimed as the consideration which should be paramount. This argument has been answered by Ronald Scott, director of planning in Greensboro, North Carolina, who points out:

> On the contrary, the people living in the Greensboro area have had every opportunity and freedom of choice and, furthermore, they have already exercised that choice. The moment that any citizen of the Greensboro area decided to choose a small plot of ground and estalish his residence thereon in close proximity to others in the urban area, his vote was cast. By the fact of his locating within this urban concentration, he has chosen to identify himself with an urban population and an urban area.

Stumbling Giants, *supra* note 65, at 65.

84. UTAH CODE ANN. §§ 10-2-101 to 104 (incorporation of city), -109 to 110 (incorporation of town), -601 to 610 (consolidation) (Supp. 1977).

85. *Id.* § 10-2-401.

86. *Id.* §§ 10-2-501 to 507.

arose recently when the Salt Lake County Commission considered a proposal to incorporate all of the unincorporated area of the county into the City of Bonneville at the same time cities within the county desired to annex areas of the unincorporated territory.[87] Conflicts between the city and the district court may also occur.[88] It is conceivable that a city could annex territory presently contiguous to the city at the same time the district court was disconnecting the contiguous incorporated territory.[89] Moreover, the Code makes no provision for one body to receive notice of conflicting proposals from another.[90]

More important than these procedural problems, however, the three separate bodies cannot coordinate their actions for the benefit of the region as a whole. At present, cities can annex areas that are highly productive sources of revenue and avoid annexation of unincorporated areas that are costly to service.[91] These decisions may be made at the expense of the county government, which has no statutory involvement in the annexation process.[92] A real need

87. Letter from Gerald H. Kinghorn, Assistant County Attorney, to Salt Lake County Board of County Commissioners (Aug. 9, 1977) (unincorporated area within the proposed limits of the City of Bonneville could not be annexed until after an election).

The recent "land war" between South Jordan and West Jordan further illustrates the problem created by conflicting jurisdictional provisions. The South Jordan City Council voted on February 21, 1978 to annex 15,000 acres, including territory West Jordan was preparing to develop. Not to be outdone, on February 28 the West Jordan City Council voted to annex a 1,500 acre strip of territory dividing the area South Jordan had annexed. The West Jordan annexation was filed with the county recorder on March 1, before South Jordan filed its annexation on March 7. Since an annexation is not complete until filing, and since state law only allows annexation of contiguous territory, the West Jordan action apparently blocked the South Jordan annexation. To further complicate the situation, Salt Lake County threatened to bring suit voiding both annexations because of the pending petition to incorporate the area into the City of Bonneville. Neither city council communicated with the other, or with the county before taking action. Deseret News, Mar. 8, 1978, § B, at 1, col. 1.

88. *See, e.g., In re* Town of West Jordan, 7 Utah 2d 391, 326 P.2d 105 (1958) (petitioners brought suit to disconnect territory that had been re-annexed two weeks after disconnection was granted in prior suit).

89. *Compare* UTAH CODE ANN. § 10-2-401 *with id.* § 10-2-501 (Supp. 1977).

90. The Code does require, however, service of a disconnection petition on the municipality. *Id.* § 10-2-501.

91. *Id.* § 10-2-401. Annexations of commercial areas or even single businesses for the obvious purpose of receiving sales tax from the point of sale, as well as property tax revenues, are numerous. For example, in Cache County, North Logan has annexed a strip of State Highway 85 north of its existing city limits and then annexed individual businesses on each side of the highway. River Heights, also in Cache County, annexed a ten-foot wide strip across a county road to reach a new subdivision. Such annexations not only affect the county's revenue base, but also upset the county's efforts in such areas as long-range planning and road improvements. Interview with G. L. Richardson, Director of Planning, Inspection, and Engineering for Cache County, in Logan, Utah (Jan. 27, 1978).

92. UTAH CODE ANN. § 10-2-401 (Supp. 1977). Interview with Gerald H. Kinghorn, Assistant Salt Lake County Attorney (Jan. 10, 1978).

exists for the input of some disinterested body that can represent area-wide concerns.[93]

The Code further prevents implementation of state policy decisions by its seemingly senseless internal inconsistency. For example, in setting out the qualifications of those entitled to sign an initiation petition, the Code creates a different class of petitioners for each of the six types of municipal boundary changes. Qualifications vary as to number, residency, ownership of property, value of property, and right to vote.[94] One wonders why property ownership, or residency, is so important in one instance and not another, or why the state requires the property owners to have one-third the value of the real property in one instance and not another. Another inconsistency has developed from recent changes in the Code. The present Utah Municipal Code seems to represent a change in state policy because it drops the requirement of the previous Code[95] that incorporation petitioners be real property owners. Yet, the Code retains such property ownership requirements for annexation and disconnection petitioners.[96] If a change in state policy was intended, the legislature failed to make it clear.

Requirements of differing petition percentages and other such distinctions could be beneficial, but only if they correspond to some general state policy. If the legislature wished to favor consolidation of cities over incorporation or annexation, for example, it could facilitate the initiation process with easier initiation requirements. The present inconsistencies do not seem to be resolved by such an explanation. Initiation requirements are easier for incorporation than for annexation, yet it is easier to obtain final approval for annexation than for incorporation.[97] Internal statutory consistency would clarify state policy, render deliberate policy decisions more effective, and make the statute easier to understand.

IV. State Interests in Municipal Boundary Changes

As subdivisions of state government, municipalities are established principally to provide services to state residents. The state,

93. California statutes require approval of the Local Agency Formation Commission (LAFCO) before any petition proposing municipal boundary changes may be circulated. Cal. Gov't Code § 35002 (West 1968). In case of conflicting proposals the commission is entitled to set the priority of consideration. *Id.* § 54798 (West Supp. 1977).

94. *See* notes 9-15, *supra* and accompanying text.

95. *Compare* ch. 55, § 1, 1899 Utah Laws 77 *with* Utah Code Ann. § 10-2-101 (Supp. 1977).

96. Utah Code Ann. §§ 10-2-401, -501 (Supp. 1977).

97. *Compare id.* §§ 10-2-101 to 108 and 10-2-401 to 404.

therefore, has a duty to see that municipalities are presently able to provide necessary services efficiently, will be able to provide such services in the future, and will not conflict with the rendering of services by other local government units. To fulfill that duty, the state legislature needs to provide a means for the proper representation of public concerns in municipal boundary changes.

The Utah Municipal Code fails to represent public concerns, but instead, seems designed to protect private interests. For example, property requirements to initiate annexation and disconnection primarily protect the property taxpayer.[98] The delegation of supervisory power over dissolution and disconnection to the district court with power to see that debts are satisfied demonstrates a desire to protect municipal creditors.[99] No comparable provisions exist to protect the public interest involved in municipal boundary changes. Final authorization of annexations[100] and town incorporations[101] is left to the complete discretion[102] of the city and county commissions respectively. A majority of voters in an election determines dissolution[103] and city incorporation[104] without any requirement that the proposal meet any area planning guidelines.

The lack of policy guidelines also leads to boundary changes motivated by the self-interest of the body authorized to render final approval.[105] This statutory deficiency has led, in the past, to annexations leaving islands of unincorporated territory surrounded by a municipality.[106] Although the legislature has since outlawed this

98. *Id.* § 10-2-401, -501.

99. The court is specifically empowered "to make provisions for the payment of all indebtedness [of the dissolved municipality] and for the performance of its contracts and obligations" and to order the levy of taxes to meet these obligations. UTAH CODE ANN. § 10-2-706 (Supp. 1977). In disconnection proceedings, the district court is granted power to ensure that obligations, incurred by the territory while part of the municipality, are satisfied.

100. *Id.* § 10-2-401.

101. *Id.* § 10-2-109.

102. See note 26 *supra*.

103. UTAH CODE ANN. § 10-2-705 (Supp. 1977).

104. *Id.* § 10-2-106.

105. Examples of incorporation action to avoid annexation or the extension of urban-type controls are numerous. Incorporations have been instigated by groups who wished to be free of county zoning and building regulations, to preempt the tax base created by the establishment of a new industry, and to provide a liquor license for the sponsors of the incorporation petition because state law would not allow such license to be granted in an unincorporated area. Cities also have been incorporated as a special haven for industry, to preserve a climate free of burdensome regulations on dairy farmers, and to assure that a community's inhabitants could continue to play draw poker without interference. S. SATO & A. VAN ALSTYNE, STATE AND LOCAL GOVERNMENT LAW 48 (2d ed. 1977) *quoting* ADVISORY COMMISSION ON INTERGOVERNMENTAL RELATIONS, GOVERNMENTAL STRUCTURE, ORGANIZATION AND PLANNING IN METROPOLITAN AREAS 39 (1961).

106. One well-known example in Salt Lake County is Sandy City, which has numerous unincorporated islands and peninsulas within its boundaries.

practice,[107] the Code still allows partial annexation of the existing islands when the city commission finds it is "in the interest of the municipality."[108] No mention is made of the interests of the inhabitants of the island area.

Oddly, disconnection is the sole municipal boundary change for which the Code specifies any public policy guidelines. The district court can order a disconnection only when "justice and equity" require it.[109] The court must also consider

> whether or not disconnection will leave the municipality with a residual area within its boundaries for which the cost, requirements, or other burdens of municipal services would become economically or practically unreasonable to administer as a municipality, . . . the effect of the disconnection on existing or projected streets or public ways, water mains and water services, sewer mains and sewer services, law enforcement, zoning and other municipal services and whether or not the disconnection will result in islands or unreasonably large or varied-shaped peninsular land masses within or projecting into the boundaries of the municipality from which the territory is to be disconnected.[110]

If the criteria have any relevance to disconnection, they have equal or more relevance to decisions about annexation, incorporation, and dissolution. The legislature ought to detail similar criteria for each proposed boundary change and provide a means for the proposal to be fully analyzed in light of such criteria.

To facilitate objective consideration of the advantages of different types of municipal boundary changes, Utah should follow the lead of several states that have established administrative commissions to supervise these changes.[111] Generally, the commissions act as disinterested bodies to assure that actions taken by municipalities comply with sound public policy as stipulated by their legisla-

---

107. Ch. 24, § 1, 1975 Utah Laws 98 (codified at UTAH CODE ANN. § 10-2-402 (Supp. 1977)).

108. UTAH CODE ANN. § 10-2-402 (Supp. 1977).

109. *Id.* § 10-2-502.

110. *Id.* § 10-2-503.

111. Though the organization, powers, and procedures to be followed by the commissions are different in each state, all such commissions review proposals for boundary changes according to guidelines established by the legislature. *See* ALASKA CONST. art. 10, § 12; ALASKA STAT. §§ 29.18.050 to .150, 29.18.250 to .460, 29.68.030 to .110, 29.68.500 to .580 (1972); CAL. GOV'T CODE §§ 54773 to 54799.5 (West 1966 & Supp. 1977); IOWA CODE ANN. §§ 368.9 to .22 (West 1976); MICH. STAT. ANN. §§ 5.2242(1)-(20) (1976); MINN. STAT. ANN. §§ 414.01 to .09 (West Supp. 1976); N.C. GEN. STAT. §§ 160A-6 to 10 (1976) (Municipal Board of Control considers only proposals for incorporation); WIS. STAT. ANN. §§ 66.013 to .022 (West 1965 & Supp. 1977) (review by both circuit court and a state planning director).

tures.[112] Some states have a commission with state-wide jurisdiction,[113] while California's Local Agency Formation Commissions (LAFCO), for example, are limited to one county.[114] Because metropolitan growth extends beyond boundary lines and the Utah Municipal Code allows annexation across county boundaries,[115] in Utah a state-wide commission would be more advantageous than commissions with jurisdiction over only one county.[116]

The powers and duties of a commission should be written clearly to identify legislative ends and to allow a commission sufficient discretion to choose appropriate means to accomplish those ends. Though some states have given their commission only non-discretionary powers,[117] discretionary powers seem preferable.[118] For example, California's LAFCO's have power to review, approve, disapprove, or alter proposals for boundary changes and the formation of special districts.[119] LAFCO's may "adopt standards and proce-

---

112. For example, California's Knox-Nisbet Act states the general policies the legislature wants to encourage:

> Among the purposes of a local agency formation commission are the discouragement of urban sprawl and the encouragement of the orderly formation and development of local governmental agencies based upon local conditions and circumstances. One of the objects of the local agency formation commission is to make studies and to obtain and furnish information which will contribute to the logical and reasonable development of local governments in each county and to shape the development of local governmental agencies so as to advantageously provide for the present and future needs of each county and its communities.

CAL. GOV'T CODE § 54774 (West 1966).

113. ALASKA CONST. art. 10, § 12; IOWA CODE ANN. § 368.9 (West 1976); MICH. STAT. ANN. §§ 5.2242(1)-(20) (1976); MINN. STAT. ANN. § 414.01 (West Supp. 1976); N.C. GEN. STAT. §§ 160A-6 to 10 (1976). The Michigan and Minnesota commissions have state-wide jurisdiction, but two of the five members are appointed on an ad hoc basis from the county involved in the proposed municipal boundary change. MICH. STAT. ANN. § 5.2242(5) (1976); MINN. STAT. ANN. § 414.01(2) (West Supp. 1977). For a description and analysis of the Minnesota Municipal Commission, see Note, *The Minnesota Municipal Commission—Statewide Administrative Review of Municipal Annexations and Incorporations*, 50 MINN. L. REV. 911 (1966); *The Minnesota Legislature, 1969 Regular Session—1969 Amendments to Minnesota's Incorporation, Consolidation, Annexation and Detachment Statute*, 54 MINN. L. REV. 1052 (1970). The director who administered Wisconsin's state-wide binary review during its early years describes that system in Johnson, *The Wisconsin Experience with State-Local Review of Municipal Incorporations, Consolidations and Annexations*, 1965 WIS. L. REV. 462.

114. CAL. GOV'T CODE § 54780 (West Supp. 1977).

115. UTAH CODE ANN. § 10-2-404 (Supp. 1977).

116. An argument also has been made that a state-wide commission would likely be staffed with higher quality personnel because "[w]ith fewer positions to fill, the appointing authority can be more selective; higher salaries will probably be paid; and statewide positions generally bestow more prestige." *Minnesota Municipal Commission, supra* note 113, at 917 n.29.

117. WIS. STAT. ANN. §§ 66.014 to .16 (West 1965 & Supp. 1977).

118. *See* Woodroof, *supra* note 18, at 754; 1969 *Amendments to Minnesota's Statute, supra* note 113 at 1057-59.

119. CAL. GOV'T CODE § 54790 (West Supp. 1977).

dures for the evaluation of proposals," "make and enforce rules and regulations for the orderly and fair conduct of hearings," and incur expenses and appoint staff or professional personnel necessary to perform these functions.[120] In addition, granting the commisson power, as Iowa has done, to initiate boundary changes, may be useful.[121] Vesting such powers in a commission does not significantly deprive the populace of power because the electorate still retains the power to vote upon the proposals.[122] The commissions act primarily as a clearing house to protect the public interest as defined by the legislature.

The legislature can regulate the discretionary power of the commission and give it direction by establishing guidelines for evaluating the merits of a municipal boundary change. Minnesota's legislature has provided one of the more detailed lists of factors for consideration by its commission.[123] The Minnesota Municipal Commission considers such factors as present population, population growth trends, the geography and usage of the land, plans for development, present governmental services, an analysis of how needed services can best be provided, fiscal data related to the proposed change, contiguity, and the effect of the change on the surrounding areas.[124]

Before adopting similar guidelines the Utah Legislature must determine what policies it desires to foster. Does it want to encourage the expansion of existing cities or the proliferation of new ones?[125] Should county governments be protected or should city-county consolidations be encouraged?[126] What measures, if any, should be taken to preserve the integrity of agricultural and rural

---

120. *Id.*

121. IOWA CODE ANN. § 368.13 (1976).

122. *Id.* § 368.19. Public hearings are also mandatory. *Id.* § 368.15.

123. Minnesota's standards have been criticized for being quite vague, despite their extensiveness. Woodroof, *supra* note 18, at 754.

124. The factors considered by the commission in all municipal boundary changes are very similar. For an example of the scope of these factors, which could serve as an excellent model for the Utah legislature, see MINN. CODE ANN. § 414.031(4) (Supp. 1976).

125. The proliferation of local government units in Utah, including special districts, has been frequently discouraged. *See, e.g.*, Van Alstyne, *Local Government Modernization: A Utah Perspective*, 1971 UTAH L. REV. 78, 78-80; Note, *Metropolitan Reorganization*, 1966 UTAH L. REV. 517; Note, *Special Districts and Deficient Local Government in the Salt Lake Metropolitan Area*, 1960 UTAH L. REV. 209.

In January 1970, a University of Utah research study recommended consolidating Salt Lake City and County governments, prohibiting the formation of new governmental units in the boundaries of the City and County of Salt Lake, and allowing only the City and County of Salt Lake to expand its boundaries. UNIVERSITY OF UTAH, LOCAL GOVERNMENT MODERNIZATION STUDY, LOCAL GOVERNMENT ARRANGEMENTS FOR MANAGING AREA-WIDE PROBLEMS, REPORT NO. 2, 49-52 (1970).

126. See recommendation in LOCAL GOVERNMENT MODERNIZATION STUDY, *supra* note 125, at 49-50.

areas? In response to policy questions such as these, the legislature should identify those factors to be considered by the commission, indicating their relative importance, that will reflect the legislature's goals. If commercially free agricultural areas are desired, for example, the commission should be required to examine "[t]he effect of the proposal on maintaining the physical and economic integrity of lands in an agricultural preserve in open-space uses."[127] If the legislature wants to protect existing local governmental units, it could require the commission to evaluate "[t]he effect of the proposed action and of alternative action on adjacent areas, on mutual social and economic interests and on the local governmental structure of the county."[128]

V. CONCLUSION

The Utah Legislature needs to make a critical and farsighted examination of the legal and social policies it wants to encourage through the Utah Municipal Code.[129] The legislature should consider extending equal access to the initiation process for municipal boundary changes by eliminating requirements of property ownership. As real property tax becomes a decreasingly important source of revenue for local governments, statutes that grant exclusive or weighted initiation power or inherent veto power to property owners become increasingly susceptible to constitutional and policy challenges.

The requirements for accomplishing the various municipal boundary changes need to be freed of jurisdictional and procedural conflicts. In addition, uniformity must be added to all phases of boundary changes. To protect public interests and to prevent the balkanization of governmental units, the legislature should adopt a state-wide commission to supervise boundary changes. With specific powers and guidelines to follow, such a commission could serve as an objective and farsighted intermediary in aiding Utah's urban growth. By failing to truly "review, modernize and codify" the law

127. CAL. GOV'T CODE § 54796(e) (West Supp. 1977).

128. *Id.* § 54796(c).

129. Before taking action on the last two parts of the Utah Municipal Code, see note 8 *supra*, the Utah Legislature should establish a committee, representative of the community, to make a study of municipal issues and recommend ways to alleviate many of the deficiencies of Chapter 2 of the Utah Municipal Code. The committee could make recommendations on such things as how citizens should participate in boundary change procedures; what differences, if any, should exist in the procedural requirements for accomplishing the various boundary changes; what the composition, jurisdiction, powers, and functional procedures of a boundary change supervisory commission should be; and, what development policies should be fostered by the Municipal Code.

dealing with municipal boundary changes, the Utah Legislature has, by default, made a decision that will perpetuate haphazard proliferation and expansion of local government units.

LYNN E. BUSATH
RANDY D. FUNK

APPENDIX

SALT LAKE COUNTY GENERAL FUND REVENUES

Revenue by Type	1972	1973	1974	1975	1976
Property Taxes	$16,932,919 58.03%	$14,687,054 38.23%	$14,217,817 37.86%	$15,747,667 32.14%	$19,797,790 31.22%
Sales Taxes	3,126,803 10.71%	2,868,752 7.47%	3,254,042 8.65%	8,559,549 17.47%	12,700,000 20.02%
Licenses & Permits	424,004 1.45%	467,853 1.22%	1,024,911 2.73%	1,152,192 2.35%	1,636,000 2.58%
Fines & Forfeitures	460,380 1.58%	515,065 1.34%	620,375 1.65%	694,186 1.42%	736,000 1.16%
Federal Grants	2,605,721 8.93%	2,453,277 6.39%	3,318,091 8.84%	5,330,426 10.88%	9,671,858 15.25%
State Grants	2,002,980 6.86%	2,356,736 6.14%	4,018,952 10.70%	4,906,430 10.01%	6,464,350 10.19%
Charges for Services	2,573,912 8.82%	2,828,226 7.36%	2,677,833 7.13%	3,208,117 6.55%	4,028,750 6.35%
Interests & Rents	612,103 2.10%	650,610 1.69%	476,026 1.27%	851,482 1.74%	1,180,000 1.86%
Other Revenues	444,581 1.52%	483,673 1.26%	418,821 1.12%	471,814 0.96%	609,000 0.96%
Federal Revenue Sharing	—0— —0—	11,098,854* 28.90%	7,528,457 20.05%	8,075,000 16.48%	6,600,000 10.41%
Total	$29,183,403 100%	$38,410,100* 100%	$37,555,325 100%	$48,996,863 100%	$63,423,748 100%

*Includes Revenue Sharing entitlements from 1972

Vertical Nonprice Restrictions and Antitrust Policy: Assumptions Re-examined

With increasing frequency scholars, courts and practitioners employ economic analysis to formulate antitrust policy.[1] Paralleling this reliance upon technical economic analysis, a growing body of doctrine has emerged dealing with the economic market impact of vertical nonprice restrictions.[2] In *Continental T. V., Inc. v. GTE Sylvania, Inc.*,[3] the Supreme Court attempted to clarify this confused and complex area of the law. The Court, relying heavily upon the economic analysis of the Chicago School,[4] held that the vertical nonprice restrictions before it were to be judged by the rule of reason[5] noting that they "allow the manufacturer to achieve certain efficiencies in the distribution of his product."[6]

In so holding, the Court overruled *United States v. Arnold, Schwinn & Co.*[7] because its "distinction between sale and nonsale transactions is essentially unrelated to any relevant economic impact."[8] The Court paid scant attention to the political and social impact of vertical restrictions on the economic freedom of the independent trader, a traditional concern of the antitrust laws.[9] The

1. Leff, *Economic Analysis of Law: Some Realism About Nominalism*, 60 VA. L. REV. 451 (1974); Polinsky, *Economic Analysis as a Potentially Defective Product: A Buyer's Guide to Posner's Economic Analysis of the Law*, 87 HARV. L. REV. 1655-56 (1974).

2. *See, e.g.*, ABA ANTITRUST SECTION, VERTICAL RESTRICTIONS LIMITING INTRABRAND COMPETITION (Monograph No. 2, 1977) [hereinafter cited as ABA Monograph No. 2]; Note, *Restricted Channels of Distribution Under the Sherman Act*, 75 HARV. L. REV. 795 (1962); Bork, *The Rule of Reason and the Per Se Concept: Price Fixing and Market Division*, Part I, 74 YALE L. J. 775 (1965); Part II, 75 YALE L.J. 373 (1966); Comanor, *Vertical Territorial and Customer Restrictions: White Motor and its Aftermath*, 81 HARV. L. REV. 1419 (1968); Posner, *Antitrust Policy and the Supreme Court: An Analysis of the Restricted Distribution, Horizontal Merger and Potential Competition Decisions*, 75 COLUM. L. REV. 282 (1975); Preston, *Restrictive Distribution Arrangements: Economic Analysis and Public Policy Standards*, 30 L. & CONTEMP. PROB. 506 (1965).

3. 97 S. Ct. 2549 (1977).

4. *See, e.g.*, R. POSNER, ANTITRUST CASES, ECONOMIC NOTES & OTHER MATERIALS (1974); Bork, *supra* note 2; Brozen, *Significance of Profit Data for Antitrust Policy*, in PUBLIC POLICY TOWARD MERGERS 110 (Weston & Peltzmann, eds. 1969); Stigler, *A Theory of Oligopoly*, 72 J. POL. ECON. 44 (1964).

5. 97 S. Ct. at 2558, 2562.

6. *Id.* at 2560. While the Court's application of the rule of reason standard to the restraints before it is technically not a substantive ruling on the status of vertical restraints in antitrust law, the Court's sweeping endorsement of the Chicago School's explanation for the legality of vertical restraints gives the practical appearance of a judicial declaration that vertical nonprice restraints are socially desireable and therefore presumptively lawful.

7. 388 U.S. 365 (1967).

8. *Id.* at 2561.

9. *See* ABA Monograph No. 2, *supra* note 2, at 25-42. Justice White in his concurring opinion had "substantial misgivings," 97 S. Ct. at 2566, for the majority's summary rejection

Court's disregard of this longstanding body of law was a necessary consequence of its heavy reliance on the Chicago School's economic analysis.[10] The Chicago School adheres to the basic tenements of the neoclassical theory of economic analysis. A fundamental proposition of the Chicago view that follows from its neoclassical assumptions is that the manufacturer imposes vertical restrictions to maximize distributional efficiencies which redound to the consumers' benefit.[11] Apparently, the Court's cursory treatment of the independent trader can be understood as a tradeoff made between the interest of the trader and the interest of the consumer. This tradeoff was prompted by the prospect of increasing distributional efficiencies which benefit the consumer. But the assumption that consumers will benefit from the imposition of vertical nonprice restrictions is open to serious question.

The purpose of this Note is to challenge the proposition, behind the Chicago School approach endorsed in GTE, that the distributional efficiencies resulting from the imposition of vertical nonprice restrictions redound to the consumers' benefit. The Note will analyze this proposition by critically examining the twin assumptions underlying neoclassical theory, that all business entities pursue maximum profits and that consumers' preferences determine society's resource allocation. It is the author's thesis that there exists a more realistically conceived first principle of firm behavior which undermines these neoclassical assumptions, thereby undercutting the proposition of the Chicago School that vertical restrictions create distributional efficiencies that benefit consumers. If it can be shown that modern firms pursue goals other than profit maximization and that these goals are not compatible with consumers' long term interests, but are nevertheless strengthened and promoted by the use of vertical restrictions, then the *GTE* Court's wholehearted endorsement of the Chicago approach to vertical restrictions is erroneous. Moreover, if it can be shown that consumers do not benefit from the division of markets caused by vertical restrictions, no justification exists for *GTE's* cursory disregard of the traditional concern for the freedom of the independent trader.[12]

---

of "this concern reflected in our interpretations of the Sherman Act for 'the autonomy of the independent businessmen.' " *Id.* at 2568.

10. *Id.* at 2558-61.

11. *See* notes 27-36 *infra* and accompanying text.

12. The purpose in pointing out the Court's disregard of the independent trader is not to defend the distinction between sale and non-sale transactions laid down in *Schwinn* nor is it to designate precisely what role the independent trader should play in the distributional chain of the economy. Rather, the purpose is to juxtapose the Court's abandonment of an entire body of case law with the Court's endorsement of the Chicago approach to vertical

I. THE ACCEPTED VIEW

A. *The Neoclassical Paradigm*

The neoclassical paradigm provides a comprehensible view of the interactions between consumers and producers. The basic system can be described as follows: Each consumer possesses preferences or tastes which relate to his consumption of various commodities.[13] The consumer expresses these tastes through his willingness to spend.[14] When the consumer exhibits a strong willingness to spend for a particular product the market price rises. This individual desire manifested through higher prices passes through the market mechanism to the producer.[15] The producer responds to this consumer willingness to spend, reflected in higher prices, by allocating resources accordingly. When producer output attains a level where consumer satisfaction derived from the last unit of expenditure for that particular product equals that for any other purpose, consumer willingness to spend becomes weak.[16] The producer stops allocating additional resources to that endeavor and reallocates them where consumer willingness to spend is strong.[17] The producer thus acts rationally by selecting a level and mix of output which maximizes his profits.[18] This system works to move resources in the

---

restrictions. This juxtaposition reveals what prompted the Court's position: the benefit to the consumer of vertical nonprice restrictions, assumed by the Chicago School, justifies the disregard of the independent trader. If it can be shown that what the Court accepted in the tradeoff between trader and consumer interests cannot be justified on the grounds of benefiting the consumer, the rationale for the Court's decision in *GTE* ceases to exist. While this situation would logically result in a resurgence of the independent trader issue, that issue is beyond the scope of this note.

13. Polinsky, *supra* note 1, at 1666.

14. R. HEILBRONER, THE ECONOMIC PROBLEM 67 (2d ed. 1970). "By a process of voluntary exchange resources are shifted to those uses in which the value to the consumer, as measured by the consumer's willingness to pay, is highest. When resources are being used where their value is greatest, we may say that they are being employed efficiently." R. POSNER, ECONOMIC ANALYSIS OF LAW 4 (1973). For Posner efficiency is "a technical term: it means exploiting economic resources in such a way that human satisfaction is measured by aggregate consumer willingness to pay for goods and services is maximized. Value too is defined by willingness to pay." *Id.* The economist makes no judgment about desires of the individual; he does not examine the source of their desire. The way the culture shapes consumer behavior is not considered. The real impact of advertising is not explored. But the cultural molding of consumer behavior is a very significant part of the modern economy and raises serious questions for antitrust law about society's direction. *See* note 84 *infra* and accompanying text.

15. *See* R. HEILBRONER, *supra* note 14, at 67; J. GALBRAITH, ECONOMICS & THE PUBLIC PURPOSE 12 (1973).

16. R. HEILBRONER, *supra* note 14, at 425-31. *See* J. GALBRAITH, *supra* note 15, at 12-13.

17. R. HEILBRONER, *supra* note 14, at 67. *See* J. GALBRAITH, *supra* note 15, at 11-13.

18. G. STIGLES, THEORY OF PRICE 149 (rev. ed. 1952); Bork Part II, *supra* note 2, at 393-95, 400-05. *See* J. GALBRAITH, *supra* note 15, at 12; J. GALBRAITH, THE NEW INDUSTRIAL STATE 123-26 (1967).

directions of consumer demand.[19] The economy attains equilibrium when supply equals demand for each factor of production and each final commodity.[20] The prices at which supply equals demand in all markets become the competitive market prices and, as such, the most efficient allocation of society's resources.[21]

Two concepts anchor the neoclassical paradigm: that the producer's primary goal is to maximize profits regardless of the firm's power to control the market[22] and that consumers' preferences determine the allocation of society's resources.[23] These concepts are corollaries of each other. The system reveals that the producer who seeks to maximize profits, regardless of firm power, will be subject to the workings of the market mechanism. This is because the market mechanism works by translating consumer preferences into producer resource allocations. If the producer seeks maximum profits, he must, by definition, either submit to or be controlled by the messages of the consumer in the form of increased or diminished purchases. When the firm seeks to maximize profits, consumer preference is "the only instruction to which the firm respond[s]. This instruction tells them where they can find the greatest possible profit, which is their sole interest."[24] Thus as "long as the economic system is imagined to be in ultimate service of the individual—to be subordinate to his needs and wishes—it can be supposed that the functions of economics is to explain the process by which the individual is served."[25] Whatever its shortcomings and predictive inaccuracies, the beauty of the neoclassical paradigm is its apparent neutrality between competing value choices.[26]

B. The Neoclassical Paradigm and the Lawfulness of Vertical Nonprice Restrictions

Robert H. Bork and Richard A. Posner, the most persuasive

19. Bork Part II, *supra* note 2, at 393.

20. Polinsky, *supra* note 1, at 1666.

21. *Id.*; R. POSNER, *supra* note 14. *See also* Bork & Bowman, *The Crisis in Antitrust*, 65 COLUM. L. REV. 363, 365 (1965).

22. Traditional economic theory holds profit maximization to be "the strongest, the most universal, and the most persistent of the forces governing entrepreneurial behavior." G. STIGLES, *supra* note 18, at 149.

23. R. HEILBRONER, *supra* note 14, at 67. *See* J. GALBRAITH, *supra* note 15, at 11-18; J. GALBRAITH, *supra* note 18, at 120-38.

24. J. GALBRAITH, *supra* note 15, at 16.

25. *Id.* at 4.

26. The neoclassical model does not purport to make moral judgments and is generally considered by its proponents to be an almost value-free system of social ordering. In reality, however, the system assumes a position of moral correctness by virtue of its unarticulated, but firmly rooted, "consumer is king" ideological foundation.

proponents for vertical nonprice restrictions, whose theories were endorsed by the Court in *GTE*,[27] argue that all such restrictions should be lawful[28] because they promote efficiency[29] in the distribution of manufactured goods. Bork and Posner conclude that such restraints are justified regardless of the consequences to the individual trader's freedom[30] on the ground that the welfare of the individual consumer should be the sole concern of antitrust law in our consumer-oriented society.[31] They argue that since firms seek to maximize profits,[32] businesses attempt either to gain monopoly power,[33] *i.e.*, control over the impersonally set market structure, or to improve manufacturing and distributional efficiencies which increase profits.[34] Thus the sole reason manufacturers impose vertical restrictions is to reduce their marginal costs of operation and thereby increase net revenues.[35] The realized cost reduction "will also, of course, increase the amount of . . . service offered the public and *lower the price.*[36] In short, if the manufacturer did not predict "that efficiency would be the net result [from which the consumer benefits] it would have no incentive to enter into the arrangement in the first place."[37]

From the fundamental assumption of profit maximization as the firm goal, it follows that any type of vertical restriction offering to increase the efficiency by which consumers receive the goods of their sovereign choice is necessarily desirable.[38] This efficiency

---

27. 97 S. Ct. at 2560-61. The Court summarily rejected the arguments of the leading critic of Bork's view, William A. Camanor. *Id.* at 2561 n.25.

28. Bork, Part II, *supra* note 2, at 397; R. POSNER, *supra* note 2, at 283-86.

29. The notion that the purpose of the antitrust laws is to promote economic efficiency runs throughout Bork's and Posner's writings. *See, e.g.*, Bork Part II, *supra* note 2, at 375-77; Bork & Bowman, *supra* note 21, at 409-10. *See generally* R. POSNER, *supra* note 14.

30. *See generally, The Goals of Antitrust: A Dialogue on Policy*, 65 COLUM. L. REV. 363 (1965) (a series of exchanges over the goals of antitrust policy between Professors Blake and Jones of Columbia and Professors Bork and Bowman of Yale).

31. Bork, Part I, *supra* note 2, at 781, Part II, *supra* note 2, at 375; Bork & Bowman, *supra* note 21, at 409-10.

32. Bork, Part II, *supra* note 2, at 393-94; Posner, *supra* note 2, at 282.

33. Bork, Part II, *supra* note 2, at 375, 393-94; Posner, *supra* note 2, at 282. *See* note 22 *supra.*

34. *See* Bork Part II, *supra* note 2, at 401. However, horizontal agreements among distributors or retailers coerced or imposed upon a manufacturer are looked upon as antisocial because their purpose is to gain market control over prices causing a restriction of output and decreased quality of goods. For an explanation in detail of this phenomenon, see Bork, Part II, *supra* note 2, at 393-94. *See also* United States v. Topco Assoc., Inc., 405 U.S. 596 (1972).

35. Bork, Part II, *supra* note 22, at 393-94. It appears the *GTE* Court was persuaded by this argument. 97 S. Ct. at 2561 n.24.

36. Bork, Part II, *supra* note 2, at 401 (emphasis added).

37. *Id.* at 402.

38. *See, e.g.*, Bork, Part II, *supra* note 2, at 430-53. From the fundamental assumption that profit maximization is the firm goal, Bork persuasively argues that vertical nonprice

argument was endorsed by the Court in *GTE*. The Court's endorsement was based on an apparent assumption that firms seek to maximize profit. The validity of this assumption, however, is open to serious question.

II. The Myth of Profit Maximization

Profit maximization assumes the producer is subject to consumer preferences through the workings of the market mechanism. To the extent that there exist firms and individual entrepreneurs in the economy incapable of substantially controlling market prices, the assumption of profit maximization is valid. Such firms exercise no control over their options; they must pursue profit to stay in business. But the assumption of perfect competition has become untenable economic theory in many sectors of the economy.[39] Economic models have changed accordingly to accommodate the tendency towards economic concentration.

Yet, while admitting that oligopolistic and monopolistic industries possess control over prices, traditional economic theorists still maintain that these organizations seek to maximize profit.[40] That

---

restrictions, while foreclosing intrabrand competition, necessarily instill competitive vigor in the interbrand market. Commentators generally claim the following benefits: (1) Manufacturers increase their depth of market penetration, ABA Monograph No. 2, *supra* note 2; Bork, Part II, *supra* note 2, at 67-68; Preston, *supra* note 2, at 511; Zimmerman, *Distribution Restrictions After Sealy and Schwinn*, 12 Antitrust Bull. 1181, 1183 (1967); (2) Divided markets encourage information exchanges between manufacturers, Bork, Part II, *supra* note 2, at 439-46; Preston, *supra* note 2, at 511; (3) The amount of interbrand nonprice competition increases, thereby stimulating and improving the quality of point of sale services, advertising, and other promotional and informational activity, ABA Monograph No. 2, *supra* note 2, at 68; Bork, Part II, *supra* note 2 at 446-51; Preston, *supra* note 2, at 511; Zimmerman, *supra* at 1183; *Restricted Channels*, *supra* note 2, at 812; (4) The manufacturer promotes economic efficiency and obtains greater accuracy in distributional effort by avoiding overlapping availability and cost of point of sales services, Bork, *supra* note 2, at 451-52; Preston, *supra* note 2, at 511; Warren, *Economics of Closed-Territory Distribution*, 2 Antitrust L. & Econ. Rev. 111, 117 (1968-69); *Restricted Channels*, *supra* note 2, at 805; (5) Manufacturers prevent destructive "free rider" and "skimming the cream" problems purportedly inherent in a distributional system devoid of vertical restrictions, ABA Monograph No. 2, *supra* note 2, at 67; Bork, Part II, *supra* note 2, at 430-38; Warren, *supra* at 117; Zimmerman, *supra*, at 1182.

39. With approximately two-thirds of the country's assets controlled by two hundred corporations, it is an indulgence in fantasy to believe that all firms in this country are locked into a market mechanism over which they exercise no control. *See Hearings Before the Senate Select Subcommittee on Small Business on the Role of Small Business in our Society*, 94th Cong., 1st Sess. (1975) (statement of John J. Flynn).

40. Bork, Part II, *supra* note 2, at 392-93. Despite the traditional adherence to profit maximization, neoclassical theorists do qualify the assumption. Robert Dorfman makes the following qualification:

> On balance, the maximization hypothesis is not as firmly grounded in the facts of life as a fundamental scientific hypothesis should be. But substantial and prolonged divergencies from the behavior it implies are rare, *particularly in industries with many*

belief is in dispute here. Unless the will to make profit, like the will for sexual expression, represents a fundamental urge of mankind, the notion that profit is the primary goal of *all* economic organizations cannot withstand critical analysis.

The assumption that firms pursue maximum profits is grounded in a fundamental belief that entreprenuers are self-interested. But the ologopolistic firm's pursuit of self-interest does not translate into a pursuit of maximum profits in the modern corporate world.[41] To posit that such firms seek self-interest through profit maximization presents a simplistic and one-dimensional model. Long term considerations such as security and stability attained through minimizing risks are also realistic goals of firm self-interest.[42] Indeed it can be argued that survival is the dominant goal of all firms to which all other goals, including extraction of a partic-

---

participants. It therefore can *still be entertained* as a sound working hypothesis.
R. DORFMAN, THE PRICE SYSTEM, 42 (1964) (emphasis added). Dorfman admits that "the profit-maximizing hypothesis works better when applied to industries composed of a large number of firms than when applied to monopolies or to industries with only a few members." *Id.* at 42.

Galbraith responds by arguing that:

> [Profit maximization] does not occur in automobiles, aluminum, rubber, synthetic fabrics, transportation, tin cans, chewing gum, chocolate, glass, soap, breakfast foods, cigarettes, most electrical goods, aircraft, tractors, computers, typewriters, most chemicals, all communications and a host of other industries . . . . The defenders of maximization are seen to be perpetrating, no doubt innocently, a rather subtle trick. Profit maximization may be assumed. But as a concession to realty the industrial system—the largest, most typical and most modern part of the economy—is excluded. The captious would be critical of any description of the social geography of the United States which, by assuming away New York, Chicago, Los Angeles and all other communities larger than Cedar Rapids, was then able to describe the country as, essentially, a small-town, front-porch community.

J. GALBRAITH, *supra* note 18, at 134.

41. Ironically the motivation behind profit maximization, the desire for money and personal aggrandizement, becomes, in the modern corporate world where management and ownership are separated, a selfless labor in other's behalf. J. GALBRAITH, *supra* note 18, at 126. Even accepting for a moment that crude self-interest is the dominant motive of mankind, it would be more logical for corporate management to pursue *sufficient* profits to insure adequate corporate earnings to guarantee the firm's security and autonomy, *id.* at 46-70, 139-68; and then to pursue other management goals such as job security, promotions, fringe benefits, executive bathrooms, and private jets, which are more in tune with management self-interest.

42. "[T]he judges decision rests upon nothing more than economic theory and is entitled so far as the truth of its assertion is concerned, to no more or less respect than the merits of the theory command . . . . We may, therefore, regard the law as substantially without internal empirical support and as resting primarily upon the validity of its reasoning." Bork & Bowman, *supra* note 21, at 403.

Replacing the fundamental assumption of firm behavior with a more realistic assumption can be challenged on the basis that the latter is not empirically proved. But the assumption that firms seek maximum profits is "substantially without internal empirical support." As such either theory rests "primarily upon the validity of its reasoning."

ular rate of profit, are pursued only to the extent they are compatible with the long term goal of survival.[43]

A. *An Analogy to the Sciences*

Proponents of the neoclassical model traditionally respond to any criticism that the model's fundamental assumptions fail to correlate with the real world by suggesting that its shortcomings do not vitiate its usefulness for giving consistent and certain results in the prediction and extrapolation of future events.[44] By analogizing to the physicist's use of unrealistic assumptions of perfect vacuums and frictionless surfaces to understand the universe, the proponents of neoclassical theory argue that assumptions of profit maximization and consumer control aid in understanding economic reality.[45] In short, translating the belief that entrepreneurs seek self-interest into the axiom that all firms seek to maximize profits provides economists with a static notion of firm behavior from which economic systems can be formulated.[46]

While this analogy to the physical sciences certainly corresponds with a static Newtonian concept of the universe in which mathematical models can plot truth,[47] perhaps the self-interest of institutionalized economic power[48] is better understood by analogy to the biological sciences.[49] The biological sciences stress organism interaction with the environment to attain the organism's ultimate goal—survival. In the long run large economic organizations evolve in accordance with the interaction between organizational structure and environmental characteristics.[50] In the process, these organiza-

---

43. Profit maximization, by definition, excludes other goals. To the extent that there are firms and individual entreprenuers in the economy incapable of substantially controlling market prices, they exercise no control over their options; they must pursue maximum profits to stay in business. For these producers maximization of profit is synonomous with survival. However, with large corporations not locked into the market mechanism survival does not demand that they maximize profits. For these firms the *rate* of profit extraction can be varied so as to be compatible with long term goals of firm survival.

44. *See* Elzinga, *The Goals of Antitrust: Other Than Competition and Efficiency, What Else Counts?*, 6 U. PA. L. REV. 1191, 1211-12 (1977).

45. *Id.*

46. *See generally,* Leff, *supra* note 1.

47. *See generally, Hearings, supra* note 39.

48. The term institutionalized economic power refers generally to that concentrated sector of the economy that controls approximately two-thirds of the nation's manufacturing assets. *See generally* R. BARBER, THE AMERICAN CORPORATION (1970); J. BLAIR, ECONOMIC CONCENTRATION (1972); P. BLUMBERG, THE MEGACORPORATION IN AMERICAN SOCIETY: THE SCOPE OF CORPORATE POWER (1975); M. MINTZ & J. COHEN, AMERICA, INC. 34-75 (1971).

49. *See* Sullivan, *Economics and More Humanistic Disciplines: What are the Sources of Wisdom for Antitrust?*, 125 U. PA. L. REV. 1214, 1232-35 (1977).

50. *Id.* at 1234 & nn. 72-74.

tions alter both themselves and their environment.[51] As such, the life processes of such organizations become "not a quest for a single optimal solution,"[52] like profit maximization, but the recognition of "a series of choices or compromises among alternative possible strategies."[53] The choice among such alternatives always reflects an effort by the organization to adapt itself to, as well as to alter, the environment to insure security and survival.

If, however, one views firms as dynamic organisms which make complex choices, one is immediately confronted with disturbing questions regarding the social value of those choices. Market forces cannot explain the behavior of a firm which regularly settles for less than a maximum return on its investments. If market forces do not account for such behavior, neither do consumer preferences, which in theory control the market. What the theory of the market does provide, however, is a model which insulates firm behavior from the responsibility for the value judgments it does make. As long as theorists postulate economic man as greedily trying to get as much as possible,

> [he] will still be subject to control by the market and ultimately, as sustained by the compulsions of avarice, by the preferences of consumers, as expressed by their purchases.[54]
>
> [I]f choice by the public is the source of power, the organizations that comprise the economic system cannot have power. If the goods that it produces or the services that it renders are frivolous or lethal or do damage to air, water, landscape or the tranquility of life, the firm is not to blame. This reflects public choice.[55]

Any notion that firms do not seek maximum profits and, therefore, are not controlled by consumer preferences, threatens to undermine the moral rectitude of neoclassical economic theory. Such notions shift the source of social power from the public to the producer. "If [a firm] takes less than the maximum, if it pursues any goal other than profit, it is assuming public responsibilities which are no part of its task"[56] and must be held accountable for these activities in

---

51. *Id.*
52. *Id.* at 1234-35.
53. *Id.*
54. J. GALBRAITH, *supra* note 18, at 121.
55. J. GALBRAITH, *supra* note 15, at 5-6.
56. J. GALBRAITH, *supra* note 18, at 124. "Few trends could so thoroughly undermine the very foundation of our free society as the acceptance by corporate officials of a social responsibility other than to make as much money for their stockholders as possible." M. FRIEDMAN, CAPITALISM & FREEDOM 133 (1962). In traditional pedagogy the private sector is assumed to pursue the somewhat apolitical value of profit maximization; political values concerning resource allocation or common direction are thought best left to the individual consumer or

political terms. Abandoning the assumption of profit maximization opens the way for a flood of new, inconvenient, and sometimes disturbing questions. Such questions, critical to an effective antitrust jurisprudence, concern the modern corporation's ability to shape society and the individual to serve its needs and the ability of antitrust law to either hold this power morally and ethically accountable or to divest the large corporations of this power altogether.

B. The Planned Economy

The work of John Kenneth Galbraith[57] supports a dynamic view of corporate behavior. He asserts that the economy has undergone an evolutionary change. Management of the mature corporation[58] eschews pursuing the goal of profit maximization in favor of pursuing the long term goals of security, stability, and maximum growth.[59] To this end management seeks the freedom from control by external forces, rather than maximized profits, to secure its long term goal of corporate survival.[60] In order to insure autonomy the mature corporation must exercise a high degree of control over the total process of supply and demand from initial procurement of raw materials through the distributional channels to the ultimate purchase by the consumer. The mature corporation finds that for effective control all activities must be carefully planned.[61] Planning then replaces the market as a mechanism for determining production and consumption and provides a method of determining what to produce at a price that insures its consumption.[62]

The reader should not envisage greedy monopolists surreptitiously conspiring in smoke-filled backrooms. Nothing should be attributed to conspiracy and little to design concerning the evolution of an institutionalized economic power capable of superseding

---

his chosen representative through the political process. Friedman clearly senses the profound political implications for society when large, powerful economic organizations pursue goals other than profits. Pursuit of other goals equates with an autonomous corporate government existing alongside our chosen government; a private government without the consent of the governed.

57. J. GALBRAITH, *supra* note 18.

58. In Galbraith's terms, the mature corporation is an organization that has adapted to advanced technology and now utilizes the large amounts of capital and comprehensive planning which this technology requires. In the process it has liberated itself from the unpredictable market mechanism. *Id.* at 71-108.

59. "A secure level of earnings and a maximum rate of growth consistent with the provision of revenues for the requisite investment are the prime goals of the [mature corporation]." *Id.* at 186. *See id.* at 176-88; J. GALBRAITH, *supra* note 15, at 92-109.

60. J. GALBRAITH, *supra* note 18, at 91-93, 177-81.

61. *Id.* at 33-44.

62. *Id.* at 33-36.

the market mechanism. Complex technology and its requirements, not the insidious machinations of greedy monopolists, is the enemy of the market.[63] Technology forces a deep and narrow specialization of labor. Only by intricately organized knowledge can complex tasks, beyond the breadth and scope of any one individual's capacities, be performed.[64] The orchestration of intricate, deeply specialized functions to perform complex tasks results in longer production cycles from drawing board to finished product.[65] More lengthy production cycles as well as increased output creates a need for investment of large amounts of capital.[66] The heavy commitment of time and money to such sophisticated endeavors tends to reduce available options. The firm becomes committed to a product long before it is marketed.[67] The need for planning arising from the inflexible commitment of time and capital required to implement modern technology is graphically revealed in the example of the Ford Mustang. The Mustang design was settled in 1962 but the automobile did not come on the market until 1964. During that period, the Ford Company inflexibly committed nine million dollars to engineering and style costs and fifty million dollars to tooling up for production.[68] The costliness of a mistake under the circumstances is readily apparent. In situations involving irreversible commitment of capital long before a product is marketed, the stakes are too high to trust the needs of the mature corporation to the contingencies of the market. Therefore the firm must plan to take into account all contingencies and such planning consists primarily of minimizing or getting rid of market influences.[69]

---

63. *Id.* at 25-29, 44-45. Galbraith's thesis has been attacked because he is claimed to advocate that modern technology requires bigness in order to be effectively implemented. His critics claim that the overly concentrated sectors of the economy are not responding to technological imperatives and that a smaller scale of concentration would promote greater efficiency. *See Hearings Before Subcommittees of the Senate Select Committee on Small Business on Planning, Regulation, and Competition*, 90th Cong., 1st Sess. (1967); M. MINTZ & J. COHEN, *supra* note 48, at 21-23, 30-31, 43-44, 47, 53-54 (1971). However, the issue of technological imperative becomes significant only after it is recognized that a highly concentrated and planned sector of the economy exists. To attack Galbraith on factual grounds, without admitting that a significant sector of the economy has effectively superceded the market misses the thrust of Galbraith's thesis. The point is that a large sector of the economy is planned and this fact is not generally accepted in the field of antitrust. An effective antitrust policy must come to terms with this fact. To dismiss Galbraith on the factual grounds that technological imperatives do not necessitate planning without ever admitting planning exists is dangerously beside the point.

64. J. GALBRAITH, *supra* note 18, at 72-74.

65. *Id.* at 25.

66. *Id.* at 26.

67. *Id.* at 26-27.

68. *Id.* at 23-24.

69. *Id.* at 37.

III. Franchise System and Vertical Nonprice Restrictions

A corporation must foresee and schedule future action and prepare for contingencies if it is to avoid major risks. It must know "what its prices will be, what its sales will be, and what its costs including labor and capital costs will be and what will be available at these costs."[70] If the corporation is subject to the market it cannot accurately and consistently predict these factors. Hence it cannot plan. Controlling market influences is essential to overall effective planning by the mature corporation. Vertical restrictions help minimize market influence at the distributional level, and thereby aid planning, by dividing up markets and eliminating intrabrand competition. However, before the relationship between planning and vertical restrictions can be fully understood, it is necessary to examine again the accepted polemics on vertical restrictions.

A. The Accepted Polemics

Vertical nonprice restrictions divide up markets and eliminate intrabrand competition. The accepted view,[71] based upon the neoclassical paradigm, assumes that manufacturers generally use such limitations on the market to increase distributional efficiency, improve point of sale services, decrease costs, and pass on the savings to the consumer. Interbrand competition is thereby stimulated to an extent that offsets the anticompetitive effects resulting from foreclosing intrabrand competition.

William E. Comanor, the leading critic of such restrictions, argues that all such restrictions should be *per se* violations of the Sherman Act.[72] In his argument he implicitly challenges the fundamental assumptions of the neoclassical paradigm. Comanor argues that manufacturers institute vertical restrictions not to increase efficiency, lower marginal costs and thereby increase profits, but to engender product differentiation[73] and gain market power. Comanor argues that market power depends not only on the number and relative sizes of firms in an industry but also on product differentia-

70. *Id.*

71. See notes 27-38 *supra* and accompanying text.

72. Comanor, *supra* note 2. The Court gave cursory and superficial treatment to this cogent argument. Continental T.V., Inc. v. GTE Sylvania, Inc., 97 S. Ct. 2561 n.25. *But see* note 87 *infra.*

73. "Simply stated, product differentiation is the degree to which developed consumer loyalty to a particular product will enable sellers to charge a premium price for that product above that charged for substitute products without losing substantial sales volume." ABA Monograph No. 2, *supra* note 2, at 62 n.250.

tion.[74] Product differentiation may be desirable when it results from "superior skill, foresight and industry."[75] But when it is a function of prestige, advertising, and promotional gimmickry Comanor argues the disproportionate and artificial allocation of resources to the distributional activity clearly increases firm market power[76] and causes consumer prices to rise.[77]

Comanor claims that vertical restrictions, in protecting the manufacturer from intrabrand price competition, allow dealers the opportunity to charge extra-competitive prices to cover the advertising and promotional costs that engender product differentiation. The product differentiation, in turn, stimulates interbrand *non-price* competition which tends to insulate manufacturers from interbrand *price* competition.[78] The resulting increase in revenue allows additional expenditures on advertising and promotional activity which further increases product differentiation and thus, in an escalating spiral, increases manufacturer market power and the cost to the consumer. Absent vertical restrictions, Comanor claims, the eroding effects of intrabrand competition would drive prices below the level capable of sustaining product differentiation and its concomitant promotional and advertising activities.[79] The result would be lower costs to consumers and less market power to manufacturers.

In 1933, E. Chamberlin in *The Theory of Monopolistic Competition* said:

> "The Differentiation of the Product" is by all odds the most difficult subject of all, and the reason is not far to seek. It contains not a technique, but a way of looking at the economic system; and changing one's economic *weltanschauung* is something very different from looking into the economics of the individual firm or adding new tools to one's kit.[80]

The concept of product differentiation changes our economic *weltanschauung* by challenging a basic assumption of the classical paradigm—that the individual consumer controls resource allocation in our economic system. This assumption is of dubious validity

---

74. Comanor, *supra* note 2, at 1423.

75. *Id.* at 1424.

76. *Id.* at 1422-32, 1436-38. In Comanor & Wilson, *Advertising Market Structure and Performance*, 49 REV. OF ECON. & STATISTICS 422 (1967), the authors demonstrate that the large firm enjoys significant advantages in the use of advertising. They find a general relationship between advertising expenditures and profits.

77. Comanor, *supra* note 2, at 1425.

78. *Id.* at 1423-27.

79. *Id.* at 1426-27.

80. E. CHAMBERLIN, THE THEORY OF MONOPOLISTIC COMPETITION 204-05 (6th ed. 1948).

when the producers the individual "commands" mold his desires in the first instance and then add the cost of that persuasion to the price of the product he buys. It becomes evident that "[t]he consumer is not sovereign if he or she is subordinate, or partly subordinate, to the will of the producer. That the economy is ultimately in the service of the consumer cannot be believed if the producer can manage the consumer, can bend him to his needs."[81]

Comanor's analysis of the effects of vertical restrictions is sound as far as it goes. When his insights concerning the effects of vertical restrictions are put in the framework of a planned economy, the full impact of vertical nonprice restrictions is more clearly seen.

B. The Planning System and Vertical Nonprice Restrictions

The mature corporation seeks to emancipate itself from the unpredictability of the market by exercising control over all phases of production. It would be absurd to seek this control over production and then leave purchases to the random fate of individual entrepreneurial skill or consumer taste.[82] Effective planning requires control over consumer behavior and control over product distribution.[83] Through the use of vertical restrictions the mature corporation gains an amount of control over the action of the independent trader tantamount to vertical integration. Further, vertical restrictions provide sufficient protection from the eroding effects of intrabrand market competition to enable the firm to maintain price levels adequate to wage a successful campaign to mold and control consumer demand.[84]

---

81. J. GALBRAITH, *supra* note 15, at 134.

82. J. GALBRAITH, *supra* note 18, at 208-42.

83. J. GALBRAITH, *supra* note 15, at 135.

84. The management of the individual through the use of advertising is a subtle phenomenon. "Most goods serve needs that are discovered to the individual not by a palpable discomfort that accompanies deprivation, but by some psychic response to their possession. The further a man is removed from physical need the more open he is to persuasion as to what he buys." J. GALBRAITH, *supra* note 18, at 211. While creating and managing consumer demand is a significant part of the organic system of the American economy, Bork does not confront the issue. He claims that "[i]t is very doubtful, to say the least, that the informative and market-entrenching aspects of advertising can be separated according to dollar amounts spent. In fact, it seems doubtful that they can be separated at all." Bork & Bowman, *supra* note 21, at 411. Posner contends that to say advertising is a waste of resources "assumes . . . that we can distinguish 'informational' and 'persuasive' advertising. . . . And it also assumes that consumers are so foolish as to be willing to pay more for advertising than its value to them in helping them make choices." Posner, *Working Paper III: Advertising and Product Differentiation*, 2 ANTITRUST L. & ECON. REV. 47, 49 (1968-69). The Court in *GTE*, apparently swayed by Bork and Posner on the matter of advertising, stated that Comanor's argument "is flawed by its necessary assumption that a large part of the promotional efforts resulting from vertical restrictions will not convey socially desirable information about product availa-

Thus, in the most concentrated and powerful sector of the economy, the assumptions of the classical paradigm, that producers seek to maximize profits and consumers control the allocation of society's resources, is replaced with the following: that producers seek to gain security and stability through effective planning and a maximum rate of growth, and that consumer demand is effective molded and controlled to insure that stability and growth. With increased economic development, power passes from the consumer to the producer in determining where and how society's resources should be allocated and the use of vertical restrictions helps preserve this shift of power.

Vertical restrictions are thus crucial to the effective planning of the mature corporation and unless the goals of planning are in concert with the present and future needs of the public, vertical restrictions which promote and preserve such goals must not be wholeheartedly endorsed as redounding to the consumers' benefit.

As we turn the corner to the twenty-first century, the values of industrial organization and planning—ever-increasing production and consumption—increasingly diverge with the real needs of the public. Voracious consumption, frivolous use of resources, waste and environmental degradation are monumental problems of increasing scope and magnitude. Vertical restrictions, in dividing up markets, allow the demand-creating activities of our economy to flourish. While their use may create short run efficiencies in effectively creating and meeting an artificially induced demand, the long term costs to the individual and society far exceed the higher prices which Comanor claims are caused by this disproportionate allocation of resources to the distributional activity. Comanor's argument, denouncing vertical restrictions because they allow extra-competitive profits, engender product differentiation, and increase market power, is meritorious and may seem a sufficient basis for holding vertical restrictions *per se* illegal. However, the consumption ideology underlying our present industrial organization, which such restrictions preserve and perpetuate, is a serious future problem[85] which cannot be solved by making vertical restrictions

bility, price, quality and services." 97 S. Ct. 2561 n.25. Thus we are told that since the precise separation of socially useful and socially wasteful expenditures on advertising cannot be determined, it should not be attempted at all. But to frame the issue in terms of when advertising expenditures spill over into the socially wasteful category ignores the relationship that exists between the mature corporation's goal of growth and stability and the need to persuade individuals to consume in a spiralling pattern of voracious consumption. In terms of the goals and needs of planning, advertising never becomes a self-cancelling and function-less expenditure of resources.

85. J. GALBRAITH, *supra* note 18, at 219.

per se illegal. Indeed, if the law were to forbid manufacturers from imposing vertical restrictions on independent traders without simultaneously attempting to control massive industrial concentration, it can be argued that firms would vertically integrate to accomplish the same ends. Since effective planning greatly minimizes firm risk, the benefits derived from planning and the control of distributional outlets may outweigh the diseconomies created by vertical integration.[86]

A policy which causes firms to integrate vertically would not be desirable. It would preclude participation by the independent trader in the economy's distributional activities[87] while failing to check the consumption ideology perpetuated by the most concentrated sector of the economy. Moreover, there are profound political implications for society when large economic organizations pursue goals other than profits. Given the tendency of the concentrated sector of the economy to plan and the resulting shift of power from the many consumers to the few producers, the present method of allocating society's resources in these sectors loses its democratic justification. Political values concerning resource allocation or common direction, values thought best left for determination to the individual consumer or his chosen representative through the political process, are now determined by large corporations shaping society to serve its needs.

It is evident that a narrow focus on the distributional chain of the economy fails to come to grips with present problems. Therefore, the only way to approach the subject of vertical restrictions is to undertake a major re-evaluation of economic theory from the perspective of the planned economy. It is not sufficient to examine the distributional activity of the economic system from a new perspective, without also examining, from this same perspective, the industrial activities leading up to distribution.

While such a task is beyond the scope of this Note, it is one that must be undertaken. The development of antitrust rules or enforcement strategies is extremely important, but a purely practical orientation to problems is short-sighted. At times the law must step back and attempt to advance the understanding of antitrust without necessarily looking for immediate solutions to problems of philosophical and social dimension.[88] Second perhaps only to constitutional

86. See Preston, *supra* note 2, at 512 for a discussion of the diseconomies involved in vertical integration.

87. *See* note 12 *supra*.

88. *See generally,* Sullivan, *supra* note 49.

law, antitrust law has great potential for affecting the social and political future of the nation. In the area of antitrust, the courts and legislators are challenged with giving effect to long term societal values and with understanding the impact each decision and each statute has on those values. Since antitrust law necessarily involves a great deal of formal economic analysis, the law can be only as effective as the analysis from which it springs. The purpose here has been to test the effectiveness of the fundamental first principle of accepted economic analysis. In doing so it is hoped that the problem of vertical restrictions can be seen in a more realistic light. Since distributional systems are only the tip of the economic iceberg, unless courts and legislators begin to change their basic assumptions of firm behavior, antitrust law will operate on an unreal perception of economic reality while the real problems of economic power will remain beyond the view and control of the law.

DAVID H. COLBY

Nondeterioration and the Protection of High Quality Waters Under Federal Water Pollution Control Law[1]

The policy of prevention of significant deterioration of high quality air, which arose from the Clean Air Act of 1970,[2] created one of the greatest controversies in federal environmental law.[3] In response to a court order,[4] the Environmental Protection Agency (EPA) promulgated regulations[5] designed to prevent the significant deterioration of air quality in those areas of the United States which have high quality air.[6] The regulations were immediately attacked from all sides, with industry and many states disputing their legality and environmentalists questioning their adequacy. The battle has been fought in the press, in legal and industrial journals, before the United States Supreme Court and, most recently, on the floor of Congress.[7]

In contrast, the nondegradation policy of the Federal Water Pollution Control Act Amendments of 1972, a "sister" statute to the

---

1. 33 U.S.C. §§ 1251-1376 (Supp. V 1975), *as amended by* Clean Water Act of 1977, Pub. L. No. 95-217, § 2, 91 Stat. 1566 (1977) [hereinafter cited as the Act]. On December 15, 1977, President Carter signed into law the Clean Water Act of 1977 which amends and renames the Federal Water Pollution Control Act Amendments of 1972, Pub. L. No. 92-500, § 2, 86 Stat. 816 (1972). These amendments do not change the major substantive provisions and policies of the Federal Water Pollution Control Act Amendments of 1972, upon which this Note is based.

2. 42 U.S.C. §§ 1857-1858a (1970 & Supp. V 1975), *as amended by* Clean Air Act Amendments of 1977, Pub. L. No. 95-95, 91 Stat. 685 (1977) (to be codified in 42 U.S.C. § 7401).

3. W. RODGERS, ENVIRONMENTAL LAW 279 (1977). For an indication of the magnitude of the controversy which has surrounded the clean air nondeterioration policy, *see generally*, Hamby, *The Clean Air Act and Significant Deterioration of Air Quality: the Continuing Controversy*, 5 ENV. AFF. 145 (1976); Note, *Review of EPA's Significant Deterioration Regulations: An Example of the Difficulties of the Agency-Court Partnership in Environmental Law*, 61 VA. L. REV. 1115 (1975); Comment, *The Nondegradation Controversy: How Clean Will Our "Clean Air" Be?*, 1974 U. ILL. L. REV. 314; Comment, *Still Up in the Air Over the Clean Air Act*, 44 U. MO. KAN. CITY L. REV. 69 (1975); Comment, *The Clean Air Act Amendments of 1977: Expedient Revisions, Noteworthy New Provisions*, 7 ENVIR. L. REP. 10182 (1977).

4. Sierra Club v. Ruckelshaus, 344 F. Supp. 253 (D.D.C. 1972), *aff'd per curiam* 4 E.R.C. 1815 (D.C. Cir. 1972), *aff'd by an equally divided court, sub nom.* Fri v. Sierra Club, 412 U.S. 541 (1973).

5. Prevention of Significant Air Quality Deterioration, 40 C.F.R. § 52 (1976). The validity of the regulations was upheld against challenges from both industry and environmentalists in Sierra Club v. EPA, 540 F.2d 1114 (D.C. Cir. 1976), *vacated and remanded* 98 S. Ct. 40 (1977) (to consider the question of mootness in light of the Clean Air Act Amendments of 1977). Congress subsequently adopted a system for prevention of significant deterioration of air quality that follows the strategy devised by the EPA. *See* Clean Air Act Amendments of 1977, Pub. L. No. 95-95, § 127(a), 91 Stat. 740 (to be codified at 42 U.S.C. §§ 7470-7479).

6. 40 C.F.R. § 52.21 (1977).

7. *See generally* notes 2-5 *supra*.

Clean Air Act,[8] has received scant attention. Unlike the response to the regulations adopted under the Clean Air Act, the implementing regulations developed by the EPA for the protection of high quality waters[9] have created scarcely a ripple of controversy among scholars, environmental groups, or industry. Perhaps this lack of attention should be seen as a tribute to the EPA, signaling that it has finally succeeded in satisfying all of the rival camps while simultaneously fulfilling a congressional mandate. Or perhaps this quietude indicates that the policy's potential has not been fully realized.

This Note will explore the nondegradation policy of the Act and the legal foundations for its implementing regulations, examine the language used by the EPA in its implementing regulations, and assess the adequacy of the protection afforded high quality waters by the regulations. The Note will conclude with suggested methods by which the EPA can better protect the remaining high quality waters of the United States.

I. BACKGROUND: THE ACT AND THE ROLE OF THE EPA

During 1970 and 1971, Congress realized that, under state leadership, the country was rapidly losing the battle against water pollution.[10] On October 18, 1972, Congress brought a powerful new weapon into the battle by enacting the Federal Water Pollution Control Act Amendments of 1972,[11] bringing to fulfillment one of the most complicated and comprehensive pieces of environmental legis-

8. The laws contain many similarities. Both are products of Senator Edmund Muskie and the Subcommittee on Air and Water Pollution of the Senate Committee on Public Works. Both provide for abatement of pollution through effluent limitations on point source polluters and both utilize ambient quality standards. Committee discussions and congressional debates frequently cross-reference the Acts. *See* notes 34-36 *infra* and accompanying text. *See also* A. KNEESE & C. SCHULTZE, POLLUTION, PRICES, AND PUBLIC POLICY 30-57 (1975).

9. 40 C.F.R. § 130.17(e) (1976).

10. The predecessor law and its amendments placed nearly the entire responsibility for prevention and control of water pollution in the hands of the states. This law relegated the federal government to a supportive role of providing financial aid and technical assistance to the states. *See* note 12 *infra*. The states had made little progress in controlling new sources of pollution, let alone reclaiming already polluted waters. While 27% of the nation's waters were considered polluted in 1970, the figure had risen to 29% in 1971. COUNCIL ON ENVIRONMENTAL QUALITY, THIRD ANNUAL REPORT 11 (1972). In its 1971 hearings, the Senate Public Works Committee reported that the battle against water pollution was being lost. Preparation of water quality standards was behind schedule, enforcement of existing standards almost nonexistent, information on pollution sources and abatement techniques was lacking, and research and development programs were underfunded and pursuing obsolete technology. S. REP. No. 414, 92d Cong., 2d Sess. 4-7 (1971), *reprinted in* [1972] U. S. CODE CONG. & AD. NEWS 3668, 3671-74 [hereinafter cited as S. REP. No. 414].

11. 33 U.S.C. §§ 1251-1376 (Supp. V 1975), *as amended by* Clean Water Act of 1977, Pub. L. No. 95-217, § 2, 91 Stat. 1566 (1977).

lation ever attempted.[12] The Act outlines no small task in its stated objective of "restor[ing] and maintain[ing] the chemical, physical, and biological integrity of the Nation's waters."[13] The objective must be achieved by accomplishing a set of national goals and policies, including the elimination, by 1985, of "discharge of pollutants into . . . navigable waters,"[14] and the achievement, wherever possible, by 1983, of "water quality which provides for the protection and propagation of fish, shellfish, and wildlife and . . . recreation in and on the water."[15]

The EPA Administrator has responsibility for implementing the Act by supervising the states in development of "comprehensive programs for preventing, reducing, or eliminating the pollution of . . . navigable waters and ground waters."[16] The Administrator may issue substantive regulations setting forth minimum standards which each state must meet in order to obtain EPA approval of its program.[17] In exercise of these powers, the Administrator promulgated regulations requiring each state to adopt and implement an antidegradation policy.[18] The Administrator required that the state policy, at a minimum, provide for protection of "existing instream water uses"[19] and for the maintenance and protection of "existing high quality waters."[20] States are allowed an exception to the latter requirement, but not to the former, based on "necessary and justifiable economic or social development."[21] This exception, however,

12. The Act is the product of nearly three years of committee work, 19 days of public hearings, 45 mark-up sessions in the Senate alone, 39 sessions of conference, and several days of floor debate. Letter from Senator Edmund Muskie to Senator Jennings Randolph (Dec. 20, 1972), *reprinted in* SENATE COMM. ON PUBLIC WORKS, 93d CONG., 1ST SESS., A LEGISLATIVE HISTORY OF THE WATER POLLUTION CONTROL ACT AMENDMENTS OF 1972, at iii (1973) [hereinafter cited as LEGISLATIVE HISTORY]. The Act consolidates the 1948 Federal Water Pollution Control Act, Pub. L. No. 845, 62 Stat. 1155 (1948) [current version at 33 U.S.C. §§ 1251-1376 (Supp. V 1975)] and four major interim amendments, and refocuses the entire approach for controlling water pollution in the United States. Ambient water quality and assimilative capacity are no longer the cornerstones of the program. Instead all polluters, regardless of the condition of the receiving water, are subject both to restrictions on their effluents and a timetable for implementation of control technology. The Act continues to rely heavily upon the states for implementation and management of programs but places the ultimate responsibility for achievement of the goals in the hands of the Federal Government. S. REP. No. 414, *supra* note 10, at 1-8, U.S. CODE CONG. & AD. NEWS, at 3668-75.

13. 33 U.S.C. § 1251(a) (Supp. V 1975).

14. *Id.* § 1251(a)(1).

15. *Id.* § 1251(a)(2).

16. *Id.* § 1252(a).

17. *Id.* § 1361(a). *See also id.* § 1313(a)(2).

18. 40 C.F.R. § 130.17(e) (1976).

19. *Id.* § 130.17(e)(1).

20. *Id.* § 130.17(e)(2).

21. *Id.*

may never be applied where the high quality waters involved "constitute an outstanding National resource."[22]

II. Foundations: The Legality of the Administrator's Action

The Act does not contain express authorization for the promulgation of regulations for the development and implementation of a national antidegradation policy. Arguably, the mandate of the Act is fulfilled when all of the nation's waters meet the Act's minimum standards, *i.e.*, when the discharge of pollutants from point sources is eliminated[23] and all waters are fishable and swimmable,[24] within the time constraints imposed by Congress. The EPA Administrator chose, however, to prescribe an antidegradation policy which must be adopted by the states as part of their programs for the protection of the nation's water resources. Central to any analysis of the regulations, therefore, is an assessment of the legality of the Administrator's action. This assessment can be made by determining whether or not authority for the promotion of this policy is fairly implied from the language of the Act, found in the underlying congressional intent, or authorized by judicial interpretation.

A. *Statutory Language*

While the Act does not explicitly provide for an antidegradation policy, there is language in the Act to support an interpretation that such a requirement does exist. The stated objective of the Act is "to restore and *maintain* the . . . integrity of the Nation's waters."[25] Moreover, the congressional policies enumerated in the Act include "prevention . . . of pollution."[26] By defining pollution as a "man-made or man-induced *alteration* of the . . . *integrity* of water," the Act arguably requires elimination of all activities that degrade water quality.[27] Further support for the nondegradation policy can

22. *Id.*

23. *See* 33 U.S.C. § 1251(a)(1) (Supp. V 1975). *Id.* § 1362(12) restricts the term "discharge of pollutant" to point source discharges.

24. *Id.* § 1251(a)(2). The goal of this section, to provide water quality "for the protection and propagation of fish, shellfish, and wildlife and . . . for recreation in and on the water" by July 1, 1983, is commonly referred to as the "1983 goal" or the "fishable and swimmable" goal.

25. *Id.* § 1251(a) (emphasis added).

26. *Id.* § 1251(b), (c) (emphasis added).

27. *Id.* § 1362(19). The language of the Senate Committee on Public Works supports this interpretation:

> The Committee has added a definition of pollution to further refine the concept of water quality measured by the natural chemical, physical and biological integrity. Maintenance of such integrity requires that any changes in the environment resulting

be found in language which governs the Administrator's review of state water quality standards. The Administrator must determine that the state standards are adequate to "*enhance* the quality of water."[28] The use of this language may be construed as an indication that Congress intended more than achievement of a minimum uniform standard of quality through a clean-up of the nation's dirty waters. Congress must also have supported the concept of nondegradation of high quality waters.

On the other hand, a number of advocates support a contrary interpretation of the statutory language or an interpretation that does not read so much into the plain language of the Act.[29] These individuals argue that the objective of restoring and maintaining the integrity of the nation's waters, when read in conjunction with the goal of achieving minimum fishable and swimmable waters, means that polluted waters, once restored, should be maintained at that minimum level of quality. In order to attain that objective, the Act directs the Administrator to assist in the preparation of programs for "preventing, reducing, or eliminating"[30] pollution. No mention is made of providing protection for high quality waters. In addition, water quality standards prepared by the states[31] are to "consist of the designated uses [of water] . . . and water quality criteria"[32] which support those uses. No mention is made of an antidegradation policy which would restrict certain uses. Indeed, the states' water quality standards must expressly take into consideration the *use* of water for "public water supplies, propagation of fish and wildlife, recreation purposes, and agricultural, industrial, and other purposes,"[33] but are not required to take into considera-

---

in a physical, chemical or biological change in a pristine water body be of a temporary nature, such that by natural processes, within a few hours, days or weeks, the aquatic ecosystem will return to a state functionally identical to the original.

S. REP. No. 414, *supra* note 10, at 76, U.S. CODE CONG. & AD. NEWS, at 3742.

28. 33 U.S.C. § 1313(c)(2) (Supp. V 1975) (emphasis added).

29. Plaintiff used this argument in Commonwealth Edison v. Train, ENVIR. L. REP. PEND. LIT. 65320 (1976), *dismissed*, 7 ENVIR. REP. (BNA) 1544 (1977). A similar approach was used in the attack on the nondegradation policy of the Clean Air Act of 1970. Sierra Club v. EPA, 540 F.2d 1114, 1120 (D.C. Cir. 1976), *vacated and remanded*, 98 S. Ct. 40 (1977) (to consider the question of mootness in light of the Clean Air Act Amendments of 1977).

30. 33 U.S.C. § 1252(a) (Supp. V 1975).

31. *Id.* § 1313 requires the states to prepare water quality standards and implementation plans for carrying out the purposes of the Act and to revise those standards every three years. The standards and revisions must be submitted to the Administrator for approval. Failure to comply with this requirement results in the promulgation of state plans and standards by the EPA Administrator. Guidelines for preparation of standards, including the requirement for preparation of an antidegradation statement, are provided in 40 C.F.R. § 130.17 (1976).

32. 33 U.S.C. § 1313(c)(2) (Supp. V 1975).

33. *Id.*

tion the *preservation* of high quality water.

The statutory language, then, is open to at least two interpretations. Looking at the language alone, both interpretations appear tenable. Therefore, in order to resolve the question of whether an antidegradation policy is legal in the area of water quality, one must look beyond the statute to other indications of congressional intent.

B. Legislative History

The legislative history of the Act does not contain an explicit discussion of an antidegradation policy. There are, however, references to the policy which support the conclusion that Congress intended to make some provision for nondegradation.[34] Senator Muskie, author of the 1972 Amendments to the Federal Water Pollution Control Act, made clear his interpretation of congressional intent on this issue: "[T]he nondegradation principle . . . [was] an *essential* element of the [Water Quality Act of 1965] and . . . has been *specifically reserved* in this bill."[35] More recently, during Senate hearings on the antidegradation policy of the Clean Air Act, Senator Muskie reiterated his opinion that the entire antidegradation concept was rooted in water pollution control law: "[T]he policy of nondegradation . . . was first articulated in Federal water pollution law in 1965."[36]

Thus, according to the sponsor of the Act, Congress clearly intended to prevent the degradation of the nation's waters. While Senator Muskie's interpretation of the Act's purpose may not be determinative of the issue, further support for his interpretation may be found in the failure of Congress to repudiate five years of administrative action which supports Muskie's view.

In 1968, Secretary of Interior Udall, convinced that the federal water pollution law mandated a nondegradation policy, issued the first official pronouncement of the policy for protecting high quality waters:

---

34. *See* S. REP. No. 414, *supra* note 10, at 76-77, U.S. CODE CONG. & AD. NEWS, at 3742; LEGISLATIVE HISTORY, *supra* note 12, at 1315, 1403 (remarks of Senator Muskie); H.R. REP. No. 911, 92d Cong., 2d Sess. 85 (1972). *See also* Hines, *A Decade of Nondegradation Policy in Congress and the Courts: An Erratic Pursuit of Clear Air and Clean Water*, 62 IOWA L. REV. 643 (1977). Dean Hines concludes that "the legislative history to support finding an implied nondegradation policy in the 1972 [Federal Water Pollution Control Act] Amendments is substantially stronger than was the comparable legislative history behind the 1970 Clean Air Act Amendments." *Id.* at 688.

35. LEGISLATIVE HISTORY, *supra* note 12, at 1315 (emphasis added).

36. *Nondegradation Policy of the Clean Air Act: Hearing Before the Subcomm. on Air and Water Pollution of the Senate Comm. on Public Works*, 93d Cong., 1st Sess. 1 (1973).

> Waters whose existing quality is better than the established standards as of the date on which such standards become effective will be maintained at their existing high quality. These and other waters of the State will not be lowered in quality unless and until it has been affirmatively demonstrated to the State water pollution control agency and the Department of Interior that such change is justifiable as a result of necessary economic or social development and will not interfere with or become injurious to any assigned uses made of, or presently possible in, such waters.[37]

Secretary Udall's articulation of the nondegradation policy was codified as a part of the official EPA regulations in 1975.[38] In the years between 1968 and 1975, the Department of Interior, and later the EPA, publicly encouraged states to develop nondegradation plans.[39] During this time, Congress made no attempt to repudiate the policy and has made no attempt to overrule the EPA regulations embodying the policy. This lack of congressional response is particularly significant because it occurred during a time when Congress enacted several major amendments to the statute on which the nondegradation policy is premised. While administrative interpretation of the statute is not conclusive of congressional intent, the failure by Congress to reject the interpretation is substantial evidence of congressional agreement with the nondegradation program.[40]

It would appear, then, that the legislative history of the Act supports the conclusion that Congress intended that existing high quality waters be preserved. However, in order to adequately assess the strength of the arguments made above, one must determine how they have fared in the courts.

---

37. Press Release, Department of Interior (Feb. 8, 1968), *reprinted in* J. DAVIES, THE POLITICS OF POLLUTION 172 (1970). In 1968, the responsibility for preparation of state water quality standards was vested in the Department of the Interior.

38. 40 C.F.R. § 130.17(e) (1976).

39. *See* 40 Fed. Reg. 55,336 (1975). In general, the response of the states has been to adopt an antidegradation policy which does little more than parrot the 1968 language of Secretary Udall. *See* ENVIR. REP. STATE WATER LAWS (BNA) 621:0101 to :0114 (1973). All of these statements have been approved by the EPA. *See* 40 C.F.R. § 120 (1976).

40. This argument was recognized in NLRB v. Bell Aerospace Co., 416 U.S. 267 (1974):

> In addition to the importance of legislative history, a court may accord great weight to the longstanding interpretation placed on a statute by an agency charged with its administration. This is especially so where Congress has re-enacted the statute without pertinent change. In these circumstances, congressional failure to revise or repeal the agency's interpretation is persuasive evidence that the interpretation is the one intended by Congress.

Id. at 274-75 (citations omitted).

A contrary position was taken in Chisholm v. FCC, 538 F.2d 349 (D.C. Cir. 1976) where it was not demonstrated that Congress had actual notice of the administrative interpretation of a statute.

C. *Judicial Interpretations*

To date, only one challenge has been made to the legality of the Administrator's antidegradation policy.[41] This litigation did not result in judicial resolution of the issue. The courts, however, have had a chance to rule on the legality of the antidegradation policy of the Clean Air Act of 1970. Because the antidegradation policies for both air and water have a common origin, and because the statutes in both areas contain similar provisions, an analysis of judicial interpretations of nondegradation and clean air will shed some light on the issue of nondegradation and high quality water.

Prior to 1972, the EPA Administrator had not required that state implementation plans for air pollution control include a provision for the prevention of deterioration of air quality so long as air quality did not fall below the national secondary ambient air standards.[42] The Sierra Club brought suit, alleging that the Administrator's failure to prevent the deterioration of high quality air was contrary to the intent of the Clean Air Act and therefore amounted to a failure to perform a nondiscretionary duty imposed by that Act.[43] The suit sought a declaratory judgment that the Clean Air Act prohibited approval of any state plan which permitted deterioration of air quality which was higher than the minimum level required by the secondary standards. Looking to the purpose of the Act, its legislative history, and the statutory language employed by Congress, the district court agreed with the Sierra Club and enjoined

41. Commonwealth Edison v. Train, No. 75-C-04127 (N.D. Ill., filed Dec. 5, 1975). The suit, filed by a group of power companies, focused on the question of the legality of the EPA antidegradation regulations. Defendant EPA, and Intervenor Natural Resources Defense Counsel, filed a motion to dismiss (June 30, 1976) on the grounds of nonjustifiability for lack of ripeness. ENVIR. L. REP. PEND. LIT. 65395-96 (1976). On February 2, 1977, the court granted the motion to dismiss, without opinion. 7 ENVIR. REP. (BNA) 1546-47 (1977).

Several factors may account for the lack of litigation. Secretary Udall's antidegradation statement required that states adopt such a policy but did not provide a methodology, or even a requirement, that the states implement the policy. While the current regulations expressly require that an antidegradation policy be implemented, 40 C.F.R. § 130.10(b)(2) (1976), revised state water quality standards are not required to be submitted for EPA approval until November 1, 1978. 40 Fed. Reg. 44,335 (1975). Indeed, the dismissal in *Commonwealth Edison v. Train* was based on the fact that the suit would not be "ripe" until EPA had acted on the state plans. Such action will probably not occur until 1979.

42. The Clean Air Act of 1970 provided for national primary and secondary ambient air standards. 42 U.S.C. § 1857(c)(4) (1970) (to be recodified at 42 U.S.C. 7409). Prevention of significant deterioration was applicable only to those regions of the country in which air quality exceeded the secondary standards. The EPA regulations required only that the state implementation plan "set forth a control strategy which shall be adequate to prevent such ambient pollution levels from exceeding such secondary standard." 40 C.F.R. § 51.12(b) (1972).

43. Sierra Club v. Ruckelshaus, 344 F. Supp. 253 (D.D.C. 1972).

approval of a state plan which permitted any significant deterioration of existing clean air.[44] The legislative intent behind, and the language and purpose of, the federal water pollution control legislation are equally supportive of a similar judicial determination that a nondegradation policy exists in federal water pollution control law.[45]

In summary, although Congress has not chosen to clearly articulate its position on the protection of high quality water, it seems clear, in light of the statutory language of the Act, its legislative history, and the rulings of the courts under the Clean Air Act, that there is a substantial basis for concluding that a mandate exists for the prevention of degradation of high quality water.[46]

III. THE REGULATIONS

A. *Language and Implementation Scheme.*

"Dual federalism" might best describe the process by which the objectives of the Act are to be achieved.[47] This process envisions a balanced federal-state effort[48] directed toward the development of solutions to problems which are national in scope. The bifurcated approach provides for national uniformity in the establishment of minimum standards and criteria,[49] but at the same time recognizes

44. *Id.* at 256. Pursuant to the court's order, the Administrator disapproved all state plans which did not include a nondegradation component, and subsequently published final regulations for prevention of significant deterioration of high quality air. *See* 37 Fed. Reg. 22,836 (1971) and 39 Fed. Reg. 42,509 (1974). The regulations were upheld against attack by both environmental and industrial concerns. *See* notes 4-5 *supra.*

45. Judge Pratt, in Sierra Club v. Ruckelshaus, 344 F. Supp. 253, 255 (D.D.C. 1972), and Judge Wright, in Sierra Club v. EPA, 540 F.2d 1114, 1124-25 (D.C.Cir. 1976), placed considerable reliance upon the statutory language of "protect and enhance" in their decisions interpreting the intent of the Clean Air Act of 1970. The current version of the Water Pollution Control Act does not contain the "protect and enhance" language of the Water Quality Act of 1965 and the Clean Air Act of 1970. The change in language to "restore and maintain" is probably not critical, however, for the avowed purpose of the Act is to strengthen prior law. *See generally* notes 10-12 *supra* and accompanying text. In addition, nothing in the legislative history of the Act indicates an intention to repeal the policy of nondegradation. The Act makes express provision for preservation of existing state water quality standards, all of which contain some sort of a nondegradation policy. 33 U.S.C. § 1313(a) (Supp. V 1975).

46. *See* Hines, *supra* note 34, at 674-75.

47. This concept is described in S. REP. No. 414, *supra* note 10, at 102-07, U.S. CODE CONG. & AD. NEWS, at 3763-68 (supplemental views of Senator James Buckley).

48. *Id.* at 8, U.S. CODE CONG. & AD. NEWS, at 3675.

49. The importance of state-to-state uniformity in environmental law has pervaded environmental legislation: "Senators will recall . . . that there are three essential elements to [the Act]: Uniformity, finality, and enforceability." LEGISLATIVE HISTORY, *supra* note 12, at 162 (remarks of Senator Muskie).

> National guidelines on prevention of significant deterioration are essential to guarantee the individual states the right to decide to maintain air quality superior to

the interests of the individual states in establishing their own priorities for planning and development within their own infrastructures.[50] Although the concept of dual federalism decentralizes the responsibility for development and implementation of the antidegradation policy, a strong federal overview must be maintained to insure uniform protection of the truly national interest in the preser-

---

minimum Federal standards. Those states which desire to retain clean air will have little hope of maintaining superior air quality without such national guidelines because:

. . .

(b) industry will pressure clean air states to relax their standards with the threat of industrial relocation in other, more permissive states.

H. REP. No. 294, 95th Cong., 1st Sess. 136 (1977), *reprinted in* [1977] U.S. CODE CONG. & AD. NEWS 2207, 2345.

National uniformity in application of policy is needed because of:

. . .

(3) the need to protect any state, region or area of the country from loss of jobs or tax revenues to other states, regions, or areas which would permit significant deterioration of air quality;

(4) the need to protect states choosing to retain clean air resources from "environmental blackmail" by industrial sources that may attempt to play one state off against another with threats to locate or relocate plants in states with weaker environmental requirements;

(5) the need to protect the vital interest of all Americans in the future air quality within certain national lands already set aside by Congress for the enjoyment of all Americans of this and future generations; and

(6) the need to protect states choosing to protect clean air resources from uncontrolled emissions from sources in other states.

Id. at 140, U.S. CODE CONG. & AD. NEWS, at 2349.

The importance of uniform national effluent standards to accomplishment of the lofty objectives of the Federal Water Pollution Control Act Amendments cannot be overemphasized. The lack of uniform standards has in the past led to lax standards, and consequent water quality deterioration, in certain areas of the country. Industries were dissuaded from locating in states with strict water quality standards. Localities downstream from permissive jurisdictions were unable to maintain desired water quality. Consequently in the 1972 Amendments to the Federal Water Pollution Control Act, Congress deliberately rejected past water pollution control strategies and adopted a uniform approach to standard-setting.

Paranteau & Tauman, *The Effluent Limitations Controversy: Will Careless Draftsmanship Foil the Objectives of the Federal Water Pollution Control Act Amendments of 1972?*, 6 ECOLOGY L.Q. 1, 57 (1976).

50. 33 U.S.C. § 1251(b) (Supp. V 1975). The 1977 Amendments to the Federal Water Pollution Control Act do not change this basic concept:

[It is the] policy of Congress that the authority of each state to allocate quantities of water within its jurisdiction should not be superseded, abrogated or otherwise impaired by this Act. It is further the policy of Congress that nothing in this Act should be construed to supersede or abrogate rights to quantities of water that have been established by any state. Federal agencies are to cooperate with state and local agencies to develop solutions to prevent, reduce and eliminate pollution in concert with programs for managing water resources.

CONFERENCE REPORT ON H.R. 3199, WATER POLLUTION CONTROL ACT AMENDMENTS OF 1977, 123 CONG. REC. H12,705 (daily ed. Dec. 6, 1977).

vation of high quality water. Uniformity in application of the policy also insures that individual states are not forced to develop their high quality waters as a result of "environmental blackmail."[51]

Responding to the need for uniform standards, the EPA has adopted national requirements for the implementation of the antidegradation policy. The current regulations require that the state's antidegradation program meet four basic requirements:[52]

(1) Where deterioration in existing water quality is allowed, such deterioration must not result in interference with or injury to existing instream water uses;[53]
(2) Where a water is determined to be of "high quality,"[54] degradation is permitted only in limited circumstances. The state must provide an underlying economic or social justification for the degradation and must demonstrate that existing instream uses are protected from injury or interference;[55]
(3) If the water qualifies as an "outstanding National resource," *no* degradation is allowed;[56] and
(4) If the degradation is due solely to the introduction of a thermal discharge, the state must protect the water only to the extent necessary to provide a habitat for a balanced, indigenous population of shellfish, fish and wildlife.[57]

Rather than concentrating on a workable number of critical

51. *See* note 49 *supra*. This need for national uniformity must be kept clearly in mind throughout the following evaluation of the antidegradation regulations. A strong antidegradation policy is certain to have significant impacts upon the future development of the water resources of the nation, particularly in those areas which include pristine watersheds. The policy will act as a limitation on the states' ability to promote increased growth and development by restricting, and in some cases prohibiting, the use of high quality waters for purposes leading to degradation of quality. The limitation will undoubtedly lead to increased pressures by business and industry on the states which, in turn, will lead to increased pressures by the states—particularly the development-oriented states—on the EPA for a liberal interpretation of the nondegradation policy. Unless the nondegradation policy is evenly applied, states will likely attempt to attract industry by offering the least restrictive pollution control standards.

52. 40 C.F.R. § 130.17(e) (1976).

53. *Id.* § 130.17(e)(1).

54. High quality waters are "waters which exceed those levels necessary to support propagation of fish, shellfish and wildlife and recreation in and on the water." *Id.* § 130.17(e)(2).

55. *Id.* The election of the state to permit degradation of high quality water can be exercised only after "full satisfaction of the intergovernmental coordination and public participation provisions of the state's continuing planning process." *Id.* These procedural requirements are set out at *id.* § 130.10(2) and are discussed at note 72 *infra*.

56. *Id.* § 130.17(e)(2). Examples of these resources are "waters of National and State parks and wildlife refuges and waters of exceptional recreational or ecological significance." *Id.*

57. *Id.* § 130.17(e)(3). This section incorporates the provisions of 33 U.S.C. § 1326 (Supp. V 1975).

pollutants, as was done in the Clean Air Act,[58] by speaking in terms of "no degradation,"[59] the Administrator has provided that all possible increments of all potential pollutants[60] are to be prohibited if they interfere with the maintenance and protection of high quality waters[61] or degrade outstanding national resource waters.[62] Thus, when a state desires to permit an activity which will have an adverse impact on high quality water, the impact must be offset by imposing a limitation on other polluting activities so that no net deterioration results.[63] However, the Administrator also incorporated a safety valve for use by the states.[64] This escape mechanism provides that states may allow some degradation of high quality waters if the degradation is the result of "necessary and justifiable" economic or social development,[65] and does not involve an outstanding resource water or interfere with or become injurious to "existing instream uses."[66]

While the requirements may, on their face, seem stringent, closer examination reveals that they contain significant loopholes. Most important, the regulations do not provide detail as to what water uses must be adopted and protected by the states, nor do they provide detail as to what quality criteria must be employed for classification of those uses. In addition, no standards are provided for the establishment of an implementation mechanism, or for the designation of outstanding national resource waters. While the imposition of these uniform requirements may be a step toward protecting states from industrial pressure and interstate competition, the requirements will not, without further elaboration, fulfill the purposes of the dual federalism approach to water quality control.

The regulations contain at least five weaknesses which erode the guarantee of uniformity in application of the nondegradation policy:

---

58. Clean Air Act Amendments of 1977, Pub. L. No. 95-95, §§ 163-166, 91 Stat. 733-39 (1977).

59. 40 C.F.R. § 130.17(e)(1), (2) (1976).

60. With the exception of thermal pollution which is controlled by 33 U.S.C. § 1326 (Supp. V 1975), and 40 C.F.R. § 130.17(e)(3) (1976).

61. 40 C.F.R. § 130.17(e)(1), (2) (1976).

62. *Id.* § 130.17(e)(2).

63. U.S. ENVIRONMENTAL PROTECTION AGENCY, GUIDELINES FOR STATE AND AREA-WIDE WATER QUALITY MANAGEMENT PROGRAM DEVELOPMENT 5-15 (1976) [hereinafter cited as EPA GUIDELINES] suggest six possible techniques whereby a state can "provide for further development" while meeting the requirements of 40 C.F.R. § 130.17 (1976).

64. The EPA is careful to point out that the antidegradation policy "should not be viewed as a 'no-growth' rule." EPA GUIDELINES, *supra* note 63, at 5-15.

65. 40 C.F.R. § 130.17(e)(2) (1976).

66. *Id.*

(1) Provision for exemptions to the nondegradation requirement for "necessary and justifiable economic or social development" and vesting the discretion for determination of those exigencies in the states;
(2) Provision for the protection of high quality water in terms of maintenance of "existing instream uses;"
(3) Failure to require the adoption of a standard set of water use classifications and to prescribe standard minimum criteria for those water uses;
(4) Failure to more clearly specify criteria for the designation of outstanding national resource waters; and
(5) Failure to establish minimum procedural requirements for implementation of the policy.

B. Analysis

1. *Exceptions for Economic or Social Development*—The objective of the Act is to "restore and *maintain* the chemical, physical, and biological *integrity* of the Nation's waters."[67] This objective implies a policy of stringent nondegradation, for even minute additions of pollutants to high quality waters will impact on the integrity of the hydrosystem involved. Interpreted narrowly, a policy of nondegradation means that the current level of water quality, as defined by its chemical, physical and biological characteristics, must be maintained. At the outset, therefore, the propriety of any exemption from a strict policy of nondegradation seems questionable, absent a relevant statutory basis for the exception.[68] Assuming that some justification for the exemption from strict nondegradation exists,[69] the mandate of Congress would seem to require that exemptions be the exception rather than the rule.

The regulations, however, fail to provide clear standards for determining what kind of "economic or social development" provides the requisite degree of necessity and justification to merit this exemption.[70] Moreover, the regulations provide that the discretion

67. 33 U.S.C. § 1251(a) (Supp. V 1975) (emphasis added).

68. *Id.* § 1312 (provides for effluent limitation exemptions for point source discharges under limited circumstances). An exemption is also provided for thermal discharges. *Id.* § 1326. The Act contains no other express statutory basis for administrative exceptions.

69. The nondegradation policy of the Clean Air Act of 1970 was interpreted by the courts to mandate prevention of only "significant" deterioration of clean air. *See* notes 4-5 *supra.*

70. 40 C.F.R. § 130.17(e)(2) (1976). Some insight into the requisite "necessary and justifiable economic or social development" can be gleaned from other EPA language. For example, in the water quality standards revision process, downgrading of a currently designated use is allowed only under limited conditions. One of these conditions occurs when "[a]pplication of effluent limitations more stringent than those required [by 33 U.S.C. § 1311 (Supp. V 1975)] . . . in order to attain the existing designated use would result in

for determination of these exigencies is vested in the states themselves.[71] Some protection from degradation for high quality waters is provided by the requirement that any decision to permit degradation be subject to "full satisfaction of the intergovernmental coordination and public participation provisions of the State's continuing planning process,"[72] and to EPA review.[73] The EPA cannot, however, rely on intergovernmental politics and public participation, particularly in a development-oriented state, to insure that adequate safeguards are established to protect high quality waters from wholesale degradation.[74]

2. *The "Existing Instream Uses" Standard*—An antidegradation policy based upon a narrow interpretation of the objective of maintaining the integrity of the nation's waters should begin by requiring the adoption of specific numerical and descriptive quality criteria, determined by the Administrator to be critical for the uniform characterization of the physical, biological, and chemical integrity of all waters of the United States. The states should then be required to establish baseline or background values for these criteria

---

substantial and widespread adverse economic and social impact." 40 C.F.R. § 130.17(c)(3)(iii) (emphasis added). Although this language is somewhat stronger than that used in the antidegradation regulations, the regulations seem to envision similar criteria for exemption. The EPA GUIDELINES, *supra* note 63, at 5-8 to 5-9, explain this exemption:

> The adverse economic and social impacts resulting specifically from imposition of the controls, and reflected in primary and secondary unemployment impacts, plant closures, changes in governmental fiscal base, and other area economic indicators, are *substantial* and *widespread* in comparison to other economic factors affecting the area's economy, to national economic conditions and fluctuations, and can be expected to persist for periods longer than provided for by adjustment payments such as unemployment compensation; and they are detectable in an area appropriate for measurement, at least as large as a county or SMSA . . . . In making a determination of substantial impact, the positive economic and social impact of enhanced water quality must be evaluated.

71. 40 C.F.R. § 130.17(e)(2) (1976).

72. *Id.*

73. 33 U.S.C. § 1313(c) (Supp. V 1975) and 40 C.F.R. § 130.17(d) (1976).

74. For example, the reaction of the western states to the original announcement of the antidegradation policy in 1968 was not positive:

> Udall's decision did not put an end to the controversy, but rather kindled the fires of opposition. Criticism came from the Western Governors' Conference, the Southern Governors' Conference, the Association of Attorneys General, and the U. S. Chamber of Commerce. Governor John Love of Colorado declared, "As we Western Governors have long realized, and as I believe Eastern Governors are more and more coming to realize, the control of the use and development of water is tantamount to absolute control of the state. For we Governors to accept such an edict and to grant such power to any agency would be no less than traitorous."

J. DAVIES, *supra* note 37, at 172-73 (1970).

Donley & Hall, *Section 208 and Section 303 Water Quality Planning and Management: Where Is It Now?*, 6 ENVIR. L. REP. 50115, 50119 (1976), also foresee a less than joyful compliance with the antidegradation regulations.

and develop and implement plans for the maintenance of the baseline determinants. Any change in the quality of water which resulted in the value of one of these specified criteria falling below or exceeding the baseline value would indicate that degradation had occurred. The state would then be required to rescind the action which caused the deterioration in quality or take other remedial steps to restore the level of quality to the baseline.

Rather than adopting this straightforward approach for assurance of national uniformity, the Administrator chose to ground his policy on the protection of "existing instream water uses."[75] The basic premise of the regulations, that "existing high quality waters . . . shall be protected and maintained,"[76] seems adequate to fulfill the mandate for nondegradation. However, by allowing some degradation and by basing the minimum protections for high quality waters on the concept of maintenance of existing instream water uses, the regulations, in fact, erode the underlying national interest in a strong, *uniform* nondegradation policy. The erosion occurs because the baseline protections are shifted away from the supervision of the Administrator and into the hands of the states with their assorted approaches to water use classification and selection of water quality criteria, and their independent procedural approaches to nondegradation.

Neither the Act, the regulations, nor the EPA's guidelines provide a clear definition of the meaning of an "existing instream water use," although the meaning of this term is crucial to the soundness of the regulations. In general, "water use" seems to be equated with the state's classification of the value of a particular body of water for "public water supplies, propagation of fish, shellfish, and wildlife, recreational purposes, and agricultural, industrial, and other purposes, and also taking into consideration their use and value for navigation."[77] The meaning of the "existing instream" aspect of the Administrator's language is particularly vague. Independent interpretations of that phrase by the states will result in varying degrees of protection for high quality waters.[78]

---

75. 40 C.F.R. § 130.17(e)(1) (1976).

76. *Id.* § 130.17(e)(2).

77. 33 U.S.C. § 1313(c)(2) (Supp. V 1975). 40 C.F.R. § 130.17(b)(2) (1976) uses similar language for designation of water use classifications in the state's water quality standards.

EPA GUIDELINES, *supra* note 63, at 2-37, -38, also equate "water uses" with "appropriate beneficial uses" which are defined in the same language as that used by the Act and the regulations. In addition, the EPA GUIDELINES explain that one of the objectives of the antidegradation policy is to insure that "existing instream beneficial water uses [are] maintained and protected." *Id.* at 5-13.

78. 40 C.F.R. § 130.17(c)(2) (1976).

The nondegradation regulations assume the existence of adequate state water quality standards which reflect the water uses currently being attained as the baseline protection of high quality waters. But if an "existing instream water use" merely reflects a state's inadequate water use classification system, the regulations do not act to further the nondegradation policy. Several definitions for "existing instream water use" might be adopted by the Administrator which would conform to the use classification designations of the states and still give the regulations substance. These definitions would require the states to identify, as part of their water quality standards, additional or higher water use values which might be affected by pollution, but which are not currently included in the water use classification system.[79] For instance, adoption of the instream use concept which has recently been added in several western states[80] to the traditionally recognized beneficial uses would promote the national interest in preservation of high quality waters where a state does not now have a high quality instream use classification. This approach would eliminate the inherent nonuniformity which results from basing the nondegradation baseline on a nonuniform system of state water use classifications.

3. *Standard Use Classifications and Minimum Criteria*—Another approach to the problem of nonuniformity would be for the Administrator to require the state to employ a *standardized*

79. Many states, particularly those in the west, recognize legitimate uses for water which do not necessarily coincide with the state's water quality standards. For instance, Utah's water quality standards currently outline five stream classifications ranging from the nearly pristine "Class A" waters to the sewer-like "Class E." ENVIR. REP. STATE WATER LAWS (BNA) 926:0541 to :0544 (1973). On the other hand, Utah's system of water rights laws does not expressly recognize preservation of aesthetics or other instream values. Recognition is given only to the use of water "for irrigation, domestic or culinary, stock watering, power or mining development or manufacturing." UTAH CODE ANN. § 73-3-8(4) (Supp. 1977). Since depletions from a water system can affect water quality as drastically as inputs, and since most, if not all, western states are facing a severe appropriations crunch, it is unclear how the failure to recognize protection of instream values as valid beneficial uses will be reconciled with the nondegradation policy.

80. Idaho, Colorado, Washington, Montana, and Oregon are the only western states which have established mechanisms for protection of instream values as valid beneficial uses. These mechanisms include withdrawal from appropriation, appropriation by state agency, reservation of water, or establishment of minimum stream flows for protection of the social value of water as an "instream use." These values include fish and wildlife propagation, recreation, or simply preservation of aesthetic values. Other states, notably New Mexico and California, have specifically refused to recognize these values in the water rights allocation process. *See* Draper, *Appropriation by the State of Minimum Flows in New Mexico Streams*, 15 NAT. RES. J. 809 (1975); Tarlock, *Recent Developments in the Recognition of Instream Uses in Western Water Law*, 1975 UTAH L. REV. 871; Welsh, *Idaho Department of Parks v. Idaho Department of Water Administration: Instream Appropriation for Recreation and Scenic Beauty*, 12 IDAHO L. REV. 263 (1976).

array of uses in their water quality standards which would reflect a spectrum of qualities ranging from backpacker drinking supplies to industrial or agricultural uses.[81] This system would narrow the degree of degradation permissible within a particular use designation. Thus, when the quality of a given body of water exceeds the minimum level necessary to satisfy a given water use under a state's current classification system, the degree of degradation permissible would be narrowed by requiring the state to apply its water quality standards revision process.[82] This standardized format should also require the adoption of specific water quality criteria for the characterization of the specified array of water uses.[83] In order to make this system a true limitation on degradation, it would also be necessary to require that, in the state's water quality standards revision process, designated water uses always reflect the highest *potential* use rather than merely the use currently being attained.[84]

---

81. Such a system would also protect important instream biotic uses and processes and avoid some of the deficiencies which exist in the current regulations. For instance, Utah's water quality standards, see note 79 *supra*, provide no specific criteria for temperature or dissolved oxygen content in its high quality "Class A" waters. Some lower uses, such as "Class C" fisheries find these criteria to be critical. Likewise, some higher uses are less sensitive to biochemical oxygen demand or suspended solids content than other theoretically lower uses.

82. The regulations governing revisions to a state's water quality standards require that "[w]here existing water quality standards specify designated water uses less than those which are presently being achieved, the state *shall* upgrade its standards to reflect the uses actually being attained." 40 C.F.R. § 130.17(c)(2) (1976) (emphasis added). Thus, by requiring each state to adopt the same standardized array of uses with narrow increments of quality separating each use, the water quality standards revision process would serve to limit a state's discretion.

83. Compare the system established by the Clean Air Act Amendments of 1977. *See* note 58 *supra* and accompanying text.

The language adopted by the EPA indicates the lack of specificity for water quality standards. The EPA requires only that water quality standards "specify *appropriate* water uses." 40 C.F.R. § 130.17(b)(2) (1976) (emphasis added). These "appropriate" uses must then be characterized by specifying "*appropriate* water quality criteria to support those water uses," *id.* § 130.17(b)(3) (emphasis added), leaving the selection process of the states. The EPA GUIDELINES, *supra* note 63, at 5-10, do little to narrow this discretion, suggesting that "numerical criteria should be stated wherever possible." The planner is then referred to U.S. ENVIRONMENTAL PROTECTION AGENCY, QUALITY CRITERIA FOR WATER (July 1976) [hereinafter cited as QUALITY CRITERIA] for selection of criteria which "will represent normally acceptable levels of water quality to support the related use." The QUALITY CRITERIA provides little help in the selection process. This document merely catalogues acceptable maximum and minimum values for support of various wildlife species, human uses, and industrial needs, for some 38 physical and chemical criteria and 15 pesticides, but does not suggest which criteria should be adopted.

The EPA GUIDELINES include non-numerical or descriptive criteria—criteria essential to recreational or aesthetic values—in water quality standards to prevent "objectionable" deposits, colors, odors, taste, or turbidity; "nuisance" debris or scum; "adverse" physiological impacts; or "undesirable" aquatic life. EPA GUIDELINES, *supra* note 63, at 5-12.

84. 40 C.F.R. § 130.17(c)(3) (1976) requires only that standards reflect uses *actually*

4. *Designation of "Outstanding National Resource" Waters* —An improved definition of "outstanding National resource" waters would further promote uniformity in the nondegradation policy. Clarification of the criteria necessary for qualification as an outstanding resource water and adoption of a clear process for initiating this classification would serve to enhance national uniformity, and would help to strengthen the procedural weaknesses in the regulations. The regulations now require that outstanding resource waters be subjected to no quality degradation and receive the highest degree of protection possible, regardless of the economic and social impacts which may result. Therefore, the designation of any water as an outstanding resource water will certainly not receive liberal application by the states, especially by the water poor states. Since the process for designation is primarily in the states' discretion,[85] it is probable that only high quality waters of such a significance as to merit recognition under a separate federal statute will be totally protected from deterioration in quality. These federal designations will serve to protect only a small number of the remaining pristine watersheds of the nation. A process should be developed for accommodating more high quality waters in this category, including less notable waters which have not been singled out for separate national attention but which have significant resource values in their own right. This protection could be accomplished by requiring—rather than suggesting—that qualified waters in National Parks and Forests and on other federal lands be initially declared outstanding resource waters. The waters could then be reclassified by the states

being attained. This language is somewhat inconsistent with Secretary Udall's pronouncement on antidegradation. See note 37 *supra* and accompanying text. The danger arises where a body of water is designated as a low quality industrial source, is concurrently being used as a higher quality stock-watering supply, and is capable of being utilized for culinary purposes, even though no such demand currently exists. Which use should the state's water quality standards reflect?

85. The regulations provide limited criteria for this designation: "such as waters of National and State parks and wildlife refuges and waters of exceptional recreational or ecological significance." 40 C.F.R. § 130.17(e)(2) (1976). While it is clear what waters are to be found within national or state parks or wildlife refuges, the waters which might qualify as "exceptional recreational or ecological" waters are not so well defined. EPA GUIDELINES, *supra* note 63, at 5-14, provides that "waters which provide a unique habitat for an identified threatened or endangered species or rivers designated under the Wild and Scenic Rivers Act" are included in this classification. However, since the substantive federal laws which provide protection for instream quality values, such as (i) the Endangered Species Act, (ii) the Marine Mammal Protection Act, (iii) the Wilderness Act, (iv) the Coastal Zone Management Act, (v) the Safe Drinking Water Act, (vi) the National Historic Preservation Act, (vii) the Wild and Scenic Rivers Act, and (viii) the organic acts and executive orders creating federal land management agencies, apply to state water quality standards anyway, the EPA GUIDELINES provide little additional definitiveness. *See* 40 C.F.R. § 130.34(a)(2) (1976).

where economic or social exigencies require, subject to veto by the responsible federal land manager. Another approach would be to establish a nomination process for inclusion in the outstanding resource category which would premit interested parties, including federal land managers, to suggest to the EPA designation of certain waters as nondegradable.

The terminology of the regulations may be given additional definition, *de facto,* by the various Regional Administrators of the EPA when they review the adequacy of state water quality standards.[86] This possibility, however, does little to establish a clearly articulated uniform national standard for use by the states in their preparation of water quality standards, by the EPA in its review of those standards, by industry in its planning activities,[87] or by the public in its oversight of the entire process.[88] This lack of standards allows the Administrator the discretion to interpret the regulations either stringently or leniently. To the extent that the regulations receive a stringent interpretation, the nation's high quality waters will be afforded a high degree of protection. A lenient interpretation affords a commensurately lower degree of protection. Furthermore, since the Act vests the responsibility for first line interpretations in the various Regional Administrators, the lack of definitive standards allows a spectrum of possible interpretations. This weakness could result in varying degrees of protection for high quality waters and could significantly reduce the potential for uniform development of national standards.

5. *Minimum Procedural Requirements*—The lack of definitive procedural safeguards for implementation of the antidegradation mechanism further weakens the existing regulatory scheme. Under the current regulations, the state's continuing planing process determines whether or not to allow degradation. The regulations only require that a decision to allow degradation be preceded

86. 40 C.F.R. § 130.17(d) (1976) delegates approval or disapproval of revisions of water quality standards to the Regional Administrators.

87. Broad administrative discretion, traditionally the bane of the environmentalists, can work both ways. "The most important single characteristic of regulation under the [Act], from the discharger's perspective, is the broad discretion vested in the Environmental Protection Agency (EPA) and the states to regulate pollution discharges." Tennille, *Federal Water Pollution Control Act Enforcement from the Discharger's Perspective: The Uses and Abuses of Discretion,* 7 ENVIR. L. REP. 50091 (1977). *See also* Commonwealth Edison v. Train, No. 75-C-04127 (N.D. Ill, filed Dec. 5, 1975), ENVIR. L. REP. PEND. LIT. 65321 (1976). The impact of unclear standards on industry's planning activities was one element of plaintiff's complaint.

88. 33 U.S.C. § 1365 (Supp. V 1975). 40 C.F.R. § 130.10(a)(1) (1976) requires public participation in development and revision of water quality standards.

by "meaningful and significant"[89] intergovernmental input and "consideration"[90] of public comments. Where a discrete decision whether or not to degrade is possible, these safeguards may be adequate to prevent unbounded development of high quality waters. Additional safeguards, such as a requirement for a decision on the record,[91] provision for designation by federal land managers,[92] or an express requirement for EPA approval of all decisions which allow degradation[93] would be helpful.

However, there are other decisions which occur independently of the formal water quality standards revision process[94] that may have significant impact on the quality of water. These decisions include the National Pollution Discharge Elimination System (NPDES) permit process,[95] and area and statewide "208 water quality management programs."[96] None of the regulations' procedural safeguards, however, apply to these decisions. While the state's continuing planning process requires intergovernmental coordination, the regulations provide no mechanism to insure that states adequately coordinate the various programs that affect water quality. The complexity of providing adequate protection for high quality waters in an efficient and uniform fashion requires that certain procedural safeguards be applied to *all* water quality related programs,[97] and that a mechanism be developed that would allow the

89. 40 C.F.R. § 130.16(c) (1976).

90. *Id.* § 105.7(a).

91. Compare the requirements of the regulations under the Clean Air Act of 1970. *Id.* § 52.21(c)(3)(ii)(d).

92. Compare *id.* § 52.21(c)(3)(iii) and (iv).

93. Compare *id.* § 52.21(c)(3)(iv).

94. See notes 96-97 *infra.*

95. 33 U.S.C. § 1342 (Supp. V 1975). Although the regulations required that any permit authorizing the discharge of a pollutant issued under this section be consistent with a state's water quality standards, coordination of these activities is not vested in a single administrative body. 40 C.F.R. § 130.32 (1976).

96. Areawide Waste Treatment Management, 33 U.S.C. § 1288 (Supp. V 1975). This section is designed to establish an agency and a plan for control of point and nonpoint source pollution. Permits issued under section 1342 of the Act must not "conflict with a plan approved pursuant to [a management plan prepared under this section]." *Id.* § 1288(e). However, there is no process to assure that this effort is coordinated with the state water quality standards revision process.

97. In Utah, for instance, water quality impacts can occur under several jurisdictions. Nonpoint source impacts are controlled by an agency of county government through such processes as the issuance of building permits, storm sewer design, zoning, and agricultural control. Point source impacts are controlled by the federal government through the issuance of discharge permits under section 1342. Appropriative rights to water are administered by a branch of the state government, the Department of Natural Resources, and somewhere in the middle of this system is the state water pollution control agency, the State Division of Health. Coordination of this decentralized system on a day-to-day basis will be an overwhelming, if not impossible, task.

EPA to monitor the states' coordination of those programs.

IV. Conclusion

The regulations for preventing degradation of the remaining high quality waters of the nation may present a unique opportunity for the EPA Administrator to protect an unspoiled environment. However, the characteristics of the regulations which provide this unique opportunity—ambiguity and lack of specificity—also provide the opportunity for weak interpretation and weak enforcement in the future. Furthermore, by vesting discretion in the states for determining the permissible degree of degradation of high quality waters, the Administrator has lost the essential element of national uniformity.

The regulations can be strengthened to provide an adequate degree of protection for high quality waters. More clearly defined procedures for public participation, intergovernmental input, federal intervention, and EPA review of decisions to designate outstanding resource waters and of decisions to allow degradation of other high quality waters, should be adopted. The discretionary decision-making powers vested in the states and the discretionary standard of review applicable to the EPA should be narrowed. The necessary limitation on discretion can be accomplished by standardizing water quality criteria and water use classifications, providing more specific criteria for determination of outstanding resource waters, and allowing designation of outstanding resource waters by federal land managers. The broad national interest in the preservation of the remaining undeveloped watersheds of the nation, truly an endangered species, should not be left defenseless before the erosive forces of unbridled administrative discretion.

Stephen Hull

Double Jeopardy—A Suggested Limit on the State's Right of Appeal in Criminal Cases

On March 22, 1977, Utah Governor Scott Matheson vetoed legislation that would have significantly increased the number of situations in which the state could appeal in criminal actions.[1] Several months later, the Utah Supreme Court dismissed a state criminal appeal because the appeal was not expressly authorized by statute.[2] These two related actions indicate the basic features of the law governing prosecutor appeals in Utah. First, the legislature has authorized state criminal appeals in only a limited number of situations by listing the specific types of rulings and orders from which the state may appeal.[3] Second, a prosecutor in Utah may seek to correct a decision which he believes is erroneous by means of a direct appeal to the state supreme court. The supreme court, however, permits appeals to be taken only when expressly allowed by statute. This Note will analyze the legal structure which governs prosecutorial appeals in Utah and will suggest a new approach for defining the state's appellate rights.

I. Prosecutorial Appeals

A. *Direct Appeal*

1. *Constitutional and Statutory Provisions*—The Utah Constitution seems to grant an unlimited right of appeal to any party from a final judgment in any action. Article VIII, Section 9 states: "From all final judgments of the district courts, there shall be a right of appeal to the Supreme Court."[4] The statutes governing state

---

1. Utah Code Ann. § 77-39-4 (1953) provides:

> An appeal may be taken by the state:
>
> (1) From a judgment of dismissal in favor of the defendant upon a motion to quash the information or indictment.
>
> (2) From an order arresting judgment.
>
> (3) From an order made after judgment affecting the substantial rights of the state.
>
> (4) From an order of the court directing the jury to find for the defendant.

H.B. 103,42d Regular Sess. (1977), vetoed by the Governor, would have added the following grounds for appeal:

> (5) From an order of the court granting a motion to dismiss.
>
> (6) From an order of the court granting a mistrial.
>
> (7) From an order of the court granting a motion to suppress evidence.

2. State v. Kelbach, 569 P.2d 1100 (Utah 1977). The court also refused to interpret the language of section 77-39-4(3) broadly so as to include a sentence in a criminal action.

3. See note 1 *supra*.

4. This section also provides for appeals from the final decision of justices of the peace in criminal cases "with such limitations and restrictions as may be provided by law." A

appeals in criminal cases, on the other hand, limit the state's appellate rights to a few specific situations. Section 77-39-4 provides that the state may appeal from a judgment of dismissal upon a motion to quash an information or indictment,[5] from an order arresting judgment,[6] from an order made after judgment affecting the substantial rights of the state, or from an order of the court directing the jury to find for the defendant.[7] The state may also appeal from an order of dismissal based on any pre-trial finding of entrapment.[8] This conflict between the broad constitutional right of appeal and the narrow statutory authorization to appeal has received varying resolutions from the Utah Supreme Court.

2. *Early Court Decisions*—Early cases suggested that state appeals were not restricted by statute. In *State v. Booth,*[9] the defen-

---

defendant may appeal any final judgment, and such an appeal transfers the case to the district court for a trial *de novo.* UTAH CODE ANN. §§ 77-57-38, -43 (1953). The prosecutor has no right of appeal from a judgment in a justice's court. Castle Dale City v. Woolley, 61 Utah 291, 212 P. 1111 (1923). The prosecutor does have a limited right to appeal on the record the final judgment of a city court in a criminal case to the district courts. UTAH CODE ANN. § 78-4-17 (1977). Appeals to the district court present distinct issues and are not within the scope of this Note.

5. A motion to quash an information may be made on the grounds that the defendant was not provided an opportunity to have a preliminary hearing, that the information did not contain the proper recitals, or that the prosecuting attorney had no authority to file it.

A motion to quash an indictment may be made on the grounds that there was an irregularity in the drawing, summoning, examination, or impanelling of the grand jury, that an unauthorized person was present during the grand jury's deliberations, that the requisite number of jurors did not concur in finding the indictment, or that the grand jury was without authority to find the indictment.

Either an information or an indictment may be quashed on the grounds that it does not charge the defendant with the commission of an offense, that the prosecutor has failed to provide a bill of particulars, that it contains a misnomer, that it contains matter which constitutes a legal justification or bar to the prosecution, or that the bill of particulars indicates a defense. UTAH CODE ANN. § 77-23-3 (1953).

6. A motion in arrest of judgment may be made on the grounds that the facts proved do not constitute a public offense or that the defendant has become insane. UTAH CODE ANN. § 77-34-1 (1953).

7. The statute is in substantially the same form as when it was first enacted. Criminal Procedures Act § 361, 1878 Utah Laws 60, 137. The only amendment made to the statute is found in ch. 132, § 1 1935 Utah Laws. This amendment rewrote subsection (1) so that it referred to "motions to quash" rather than to "demurrers."

Most states have similar restrictions on the state's right of appeal. These statutory limitations have been explained as the result of three influences. First, there was the unquestioned right of appeals by the Crown at common law. Second, at the time most American statutes were passed, criminal defendants suffered under many procedural disadvantages in the courts, and the limitation on the state's right of appeal was an attempt to equalize the sides. Third, early American legislatures generally emphasized the concept of liberty, freedom, and individual rights and thus attempted to protect the defendant from abuse by the state. *See* Busch & Becker, *(Appellate) Power to the People—A Primer of Prosecution Appeals in California,* 7 U.W.L.A.L. REV. 8, 17 (1975).

8. UTAH CODE ANN. § 76-2-303(5) (1977).

9. 21 Utah 88, 59 P. 553 (1899).

dant had obtained a judgment of dismissal on the grounds that the trial court lacked jurisdiction to try him. The state petitioned the Utah Supreme Court for a writ of mandamus to compel the trial court to reinstate the case, and the court sustained a demurrer on the grounds that the state had an adequate remedy via direct appeal. The court stated:

> In Sec. 9, Art. 8, Const., it is provided: "From all final judgments of the district courts, there shall be a right of appeal to the supreme court." Here is a plain and express provision of the fundamental law, which grants the right of appeal "from all final judgments of the district courts." It is mandatory, and applies alike to criminal prosecutions and civil actions. It is a limitation alike upon the legislative and judicial powers of the government. Neither the legislature by legislation nor the judiciary by interpretation can lawfully deprive any person, natural or artificial, from this sovereign right. The state is not made an exception.[10]

After *Booth,* the court allowed a few state criminal appeals in cases which appear to be outside of the terms of the statutes which authorize appeal.[11] These cases typically involved an appeal from a judgment of dismissal in favor of the defendant. The cases do not discuss the state's right to appeal, nor do they indicate whether the decisions are based on the statutes authorizing an appeal or on the constitutional provision. None of these cases have been expressly overruled, but they are clearly out of step with the more modern decisions which have restricted the prosecutor's right to a direct appeal.

3. *Modern Authority*—Recent cases hold that the state may

---

10. *Id.* at 91, 59 P. at 554.

11. In State v. Rickenburg, 58 Utah 270, 198 P. 767 (1921), the court heard an appeal from a judgment of dismissal following a plea of not guilty. This plea is inconsistent with the demurrer. *See* Criminal Procedure Act, § 190, 1878 Utah Laws 60, 101. The motion to dismiss was based on the failure of the information to state facts constituting an offense, however, and this is a recognized ground for a demurrer. Criminal Procedure Act, § 192, 1878 Utah Laws. *See also* notes 5-7 *supra.* In State v. Thatcher, 108 Utah 63, 157 P.2d 258 (1945), the state was allowed to appeal from a judgment of dismissal entered at the close of the state's case on the grounds that the defendant had been once in jeopardy. In State v. Iverson, 10 Utah 2d 171, 350 P.2d 152 (1960), the state was allowed to appeal a judgment of dismissal entered after the judge had discharged the jury for failure to reach a verdict. Justice Henriod dissented and stated that the appeal should be dismissed for its failure to fit into an appropriate statutory "slot." *Id.* at 173, 350 P.2d at 153. Justice Henroid's dissent in this case may be properly regarded as the genesis of the modern rule restricting the state's appellate rights. *See* Boyce & Dewsnup, *Survey of Utah Law—1960,* 7 UTAH L. REV. 342, 346 (1961).

In State v. Brennan, 13 Utah 2d 195, 196 n.1, 371 P.2d 27,28 n.1 (1962), the court in a footnote stated that the appeal in the case was authorized by the section of the statute which allows the state to appeal orders directing verdict for the defendant, but the main opinion recited that the appeal is from an order of dismissal.

not appeal without express statutory authority to do so. In *Hartman v. Weggeland,*[12] for example, criminal defendants in a city court had successfully petitioned a district court for an order compelling the city court judge and the prosecutor to make certain depositions available to the defense.[13] The state sought to appeal the district court's order, and the appeal was resisted on the ground that it was an unauthorized state appeal in a criminal case. Rather than argue that its appellate rights were not restricted in a criminal case, the state sought to avoid a dismissal by asserting that the appeal was purely civil in nature. The court responded that: "The hairy hands of Esau do not fool us that we fail to detect the voice and person of Jacob," and dismissed the appeal.[14]

After *Hartman,* the court dismissed two state appeals which were not within the terms of any statute authorizing an appeal.[15] The opinions do not reveal whether the dismissals were considered mandatory or discretionary with the court. That question was put to rest by *State v. Davenport,*[16] where the majority dismissed a state criminal appeal because "[t]he state has no standing as a litigant-appellant in this case since the basis for its appeal appears to be a stranger to the only four bases upon which the State may appeal."[17] Two justices filed a vigorous dissent and argued that the majority's opinion was inconsistent with the broad grant of appellate rights found in Article VIII of the Utah Constitution.[18]

---

12. 19 Utah 2d 229, 429 P.2d 978 (1967).

13. The district courts have supervisory authority over the city courts. UTAH CONST. art. VIII, § 7; UTAH CODE ANN. § 78-3-9 (1977). The record in the case did not disclose "the who, when, where or why" of the depositions. 19 Utah 2d at 231 n.1, 429 P.2d at 979 n.1.

14. *Id.* at 230, 429 P.2d at 979. The actual holding in this case is subject to question. *See, e.g.,* Van Dam v. Morris, No. 15059 (Utah Nov. 3, 1977). In *Morris,* the supreme court allowed the state to appeal a district court denial of a petition for a writ of mandamus compelling a city court judge to set a criminal matter for a preliminary hearing. Also, the state has long enjoyed the right of appeal from orders discharging a prisoner on a writ of habeas corpus, even though such actions are collateral attacks on criminal judgments. *See, e.g.,* Winnovich v. Emery, 33 Utah 345, 93 P. 988 (1908) (overruled a line of cases holding discharge orders unappealable). The holding in *Winnovich* is based, *inter alia,* on UTAH CONST. art. VIII, § 9.

Even though the holding in *Hartman* is open to question, the case is important as the first case where the majority of the court questioned the ability of the state to appeal in a criminal case.

15. State v. Callahan, 26 Utah 2d 304, 488 P.2d 1048 (1971) (charge of resisting arrest dismissed because the statute under which the arrest was made was void); State v. Overson, 26 Utah 2d 313, 489 P.2d 110 (1971) (charge dismissed because the trial court suppressed the defendant's admission).

16. 30 Utah 2d 298, 51 P.2d 544 (1973). The order of dismissal was based on speedy trial grounds.

17. *Id.* at 299, 517 P.2d at 545.

18. *Id.* at 301, 517 P.2d at 546 (Crockett, J., dissenting). The Supreme Court of Idaho exercises a discretionary power to hear appeals by the state in criminal cases which are not

Recently, the court refused to accept the state's invitation to overrule *Davenport*.[19] Consequently the rule is now well established that a prosecutor cannot appeal outside of the four corners of the authorizing statute. No case has attempted to answer the argument of the *Davenport* dissenters, but *Castle Dale City v. Woolley*[20] provides some insight into the courts reasoning. In *Castle Dale,* the Utah Supreme Court held that a city may not appeal an adverse criminal decision in a justice's court to a district court. The Utah Constitution does not guarantee a right of appeal to the district courts,[21] nor does any statute grant a city the right to appeal in a criminal case. The court refused to imply a right of appeal and cited *United States v. Sanges*[22] in support of its conclusion. In *Sanges,* The United States Supreme Court held that a general statute granting the right to a writ of error in the federal courts did not authorize the United States to seek a writ of error in a criminal case. The Court reasoned that a general *statute* was not intended to overturn the common law rule that the sovereign has no right of appeal in a criminal case. If the court in *Davenport* was relying on the *Sanges* rationale to limit the appellate rights that are seemingly granted by the Utah Constitution, then it should have done so expressly and it should have explained why such reasoning would apply to a general *constitutional* grant of appellate rights.

B. *Extraordinary Writs*

The Utah Supreme Court has original jurisdiction to issue extraordinary writs and can issue all writs necessary and proper to the exercise of its appellate jurisdiction.[23] Special forms of pleading relating to writs have been abolished by Rule 65B of the Utah Rules

---

authorized by statute. *See, e.g.,* State v. Maddock, 97 Idaho 610, 549 P.2d 269 (1976).

19. State v. Kelbach, 569 P.2d 1100 (Utah 1977). The defendants had been sentenced to death prior to Furman v. Georgia, 408 U.S. 238 (1972), and their sentences had been vacated by the United States Supreme Court. Kelbach v. Utah, 408 U.S. 935 (1972). The Utah Supreme Court remanded the case to the trial court for further proceedings. State v. Lance, 559 P.2d 543 (Utah 1977). The Utah Attorney General asked that a jury be convened to consider the imposition of a death sentence, but the court imposed life sentences. The state appealed from this order, and the defendants moved to dismiss the appeal. In addition to arguing that *Davenport* should be overruled, the state contended that the sentences were orders made "after judgment affecting the substantial rights of the state" and therefore appealable. *See* People v. Gilbert, 25 Cal. 2d 422, 154 P.2d 657 (1944); State v. Alexander, 15 Utah 2d 14, 386 P.2d 411 (1963). The court, nevertheless, refused to hold that a sentence is an order made after judgment.

20. 61 Utah 291, 212 P. 1111 (1923).

21. See note 4 *supra.*

22. 144 U.S. 319 (1892).

23. UTAH CONST. art. VIII, § 4, UTAH CODE ANN. § 78-2-2 (1953).

of Civil Procedure; the rule does provide, however, that relief may be obtained where an appeal or other adequate remedy is unavailable or where an inferior tribunal has exceeded its jurisdiction, abused its discretion, or failed to perform a duty specifically enjoined by law. The relief may take the form of an order reversing or vacating the judgment rendered in the trial court, or it may consist of an order compelling the trial court to perform a specific duty. The principal differences between review by appeal and by writ are that a writ may only reach those grounds of error specified in the rule whereas an appeal can review any error affecting the judgment,[24] and the court will not grant a writ as a matter of right.[25]

Although the state has rarely used them, extraordinary writs may be an important means of obtaining appellate review for the state in criminal cases. The Utah Supreme Court has issued writs on the state's petition to compel a judge to impanel a twelve member jury in a homicide case,[26] to reverse an order quashing an information,[27] and to reverse a district court's dismissal of an appeal from a city court.[28] No case has indicated that the state's right to seek the writs differs from the right of any other litigant, and no statute purports to limit the state's right to seek review by means of extraordinary writs.[29]

There are relatively few cases in which the state has petitioned the supreme court for an extraordinary writ in a criminal matter. In view of the state's qualified right to appeal criminal cases, the failure of the state to make more frequent use of the writs to obtain review is surprising. The grounds necessary for such writs, especially "excess of jurisdiction" and "abuse of discretion," are fairly elastic

---

24. *Compare* UTAH CODE ANN. § 77-42-1 (1953) *with* UTAH R. CIV. P. 65B.

25. L. J. Mueller Furnace Co. v. Crockett, 63 Utah 479, 227 P. 270 (1924).

26. State v. Hart, 19 Utah 438, 57 P. 415 (1899).

27. Higgins v. Burton, 64 Utah 550, 232 P. 917 (1925). The order was considered unappealable because no judgment of dismissal had been entered. *See, e.g.*, State v. Thompson, 69 Utah 282, 254 P. 147 (1927). This procedural nicety is no longer important, and the supreme court will allow the state to appeal directly from an order quashing an information without requiring a judgment of dismissal. State v. Ward, No. 14903 (Utah Nov. 9, 1977).

28. Salt Lake City v. Hanson, 19 Utah 2d 32, 425 P.2d 773 (1967). In State v. Ruggeri, 19 Utah 2d 216, 429 P.2d 969 (1967), the state sought review of an order suppressing evidence that had been obtained during a grand jury investigation on the grounds that its use during trial would violate the defendant's right against self-incrimination. Although the main opinion in the case indicates that an extraordinary writ was unavailable to review evidentiary rulings, a majority of the court affimed the lower court's suppression order on the merits.

29. In Higgins v. Burton, 64 Utah 550, 553, 232 P. 917, 918 (1925), the court indicated that the writ was available because the state had no right of appeal. Other jurisdictions, however, have held that a statute denying the state a right of appeal also restricts the state's right to seek an extraordinary writ. *See, e.g.*, People v. Superior Court, 69 Cal. 2d 491, 72 Cal. Rptr. 330, 446 P.2d 138 (1968); Annot., 91 A.L.R.2d 1095 (1963).

and could be used to seek review of a wide variety of claimed errors.[30] Because of their availability and the scope of review they offer, the extraordinary writs could be used to compensate for the prosecutor's inability to appeal in many cases.

C. *The Impact of the Double Jeopardy Clause*

Article I, Section 12 of the Utah Constitution provides that no person "shall . . . be twice put in jeopardy for the same offense."[31] This constitutional clause protects a criminal defendant from the harassment, uncertainty, expense, and delay of a second trial and forbids a second punishment for the same crime. Thus, even should the state convince the supreme court that a decision in the trial court was erroneous, it would be denied a remand for a new trial in those cases where a new trial would constitute a second jeopardy. This double jeopardy limitation on the state's ability to obtain relief on appeal caused confusion in early cases.[32] But later cases have established that an appellate reversal of a judgment favorable to a criminal defendant does not result in any further proceedings if jeopardy "attached" during the proceedings below.[33] In Utah, as in most other jurisdictions, jeopardy attaches when the jury is impanelled or sworn, or, in the case of a bench trial, when the first witness is sworn.[34] Thus only if the judge dismisses the case before jeopardy has attached, can the defendant be compelled to stand trial if the dismissal is reversed on appeal.[35]

One unresolved question in Utah is whether a state appeal from an order made after judgment would have the effect of reinstating the judgment. This issue would arise, for example, in a state appeal of an order arresting judgment. Such an appeal would clearly be

---

30. It thus appears that even though an absolute lack or excess of jurisdiction cannot be shown, the writ can be and, in practice, is issued in the sound discretion of the higher court. . . . [I]t would be impossible to formulate . . . any statement by which an excess of jurisdiction could be definitely determined and distinguished from mere error. Olson v. District Court, 93 Utah 145, 157, 71 P.2d 529, 534 (1937).

31. This constitutional protection is also provided by UTAH CODE ANN. § 77-1-10 (1953). The parallel federal right is guaranteed by U.S. CONST. amend. V, which is made applicable to the states by the due process clause. Benton v. Maryland, 395 U.S. 784, 793-796 (1969).

32. *See* State v. Gustaldi, 41 Utah 63, 72, 123 P. 897, 900-901 (1912).

33. State v. Brennan, 13 Utah 2d 195, 371 P.2d 27 (1962); State v. Iverson, 10 Utah 2d 171, 350 P.2d 152 (1960); State v. Sandman, 4 Utah 2d 69, 286 P.2d 1060 (1955); State v. Thatcher, 108 Utah 63, 157 P.2d 258 (1945); State v. Cheeseman, 63 Utah 138, 223 P. 762 (1924).

34. *See, e.g.*, State v. Whitman, 93 Utah 557, 74 P.2d 696 (1937); UTAH CODE ANN. § 76-1-403(4) (1977).

35. State v. Bridwell, 556 P.2d 1232 (Utah 1977); State v. Conover, 28 Utah 2d 335, 502 P.2d 552 (1972).

made after jeopardy has attached, but would require no retrial or further proceedings. Presumably, the Utah courts would follow the United States Supreme Court's analysis in *United States v. Wilson*[36] and allow an appeal from an order made after a conviction to result in the reinstatement of the guilty verdict. This procedure would not offend the double jeopardy clause because the verdict could be reinstated by the reviewing court without subjecting the defendant to further proceedings before the judge or jury, i.e., without putting the defendant in jeopardy again.

II. Analysis of the Prosecutor's Appellate Rights

A. *Circumstances in Which the State's Criminal Appellate Rights Are Too Broad*

The present Utah law permits the state to appeal in cases where the judgment on appeal cannot affect the defendant because he is insulated from further proceedings by the constitutional protection against double jeopardy.[37] Because this kind of an appeal cannot result in a judgment which affects the parties, it must be regarded as a request for an advisory opinion.[38] These advisory opinions are objectionable for two reasons. First, they are a wasteful use of scarce judicial resources. Second, they can produce bad law because they are not decided in an adversarial context. The opinions tend to be lopsided in favor of the prosecution because the defendant has no incentive to contest the state's points on appeal.[39]

The problems involved in an advisory opinion were set forth in Chief Justice Larson's dissent in *State v. Thatcher:*

> Appeals by the state in criminal cases lie only on questions of law, since the defendant cannot be again brought to trial. An appeal therefore that does not settle a point of law which will be helpful in future cases is wholly abortive, a waste of time, effort and expense. This appeal settles nothing except that the prosecutor can say to the judge, now off the bench, "I told you so." The opinion is no guide or help in

36. 420 U.S. 332 (1975).

37. This problem is alleviated to a degree by the rule of *Davenport,* because appeals of most dismissals are now prohibited. The double jeopardy limitation would still be a factor in appeals from directed verdicts and from dismissals upon motions to quash entered after jeopardy has attached.

38. An advisory opinion characteristically has no precedential value and involves no judgment against a party. *See* Note, *Judical Determinations in Nonadversarial Proceedings,* 72 Harv. L. Rev. 723, 732 (1959).

39. In State v. Davenport, 30 Utah 2d 298, 517 P.2d 544 (1973), the defendant did not resist the state's attempt to obtain review. In State v. Cheeseman, 63 Utah 138, 223 P. 762 (1924), the defendant did not file a brief on appeal.

> future cases. . . . Furthermore Section 105-43-1, U.C.A. 1943[40] provides that this court must give judgment without regard to errors unless satisfied that but for such error the judgment may well have been otherwise, and that "it shall not be presumed to have resulted in prejudice." Since the effect of the judgment, and the position and rights of the parties are the same regardless of reversal, and no law question is settled for future cases, the whole thing is in the nature of a sideshow—a moot entertainment without effect.[41]

Because the appeal of moot cases cannot further any legitimate governmental interest, the statutes authorizing the state to take such appeals should be reconsidered.

B. *Circumstances in Which the State's Criminal Appellate Rights Are Too Narrow*

Current Utah law prevents the state from seeking appellate review in cases where the review could serve to vindicate important public policies. For example, a guilty defendant may escape punishment because of a trial court's erroneous dismissal of an information or indictment on the grounds of an allegedly void statute, denial of the right to a speedy trial, or double jeopardy.[42] The state's inability to appeal in these cases impairs several legitimate governmental interests.

The state has an interest in maintaining public respect for the criminal justice system.[43] When an offender escapes an otherwise appropriate criminal sanction because of a judge's unreviewable error of law, the public becomes suspicious and disrespectful of the justice system.

Also, state appeals can further the development of a uniform body of criminal law.[44] This interest encompasses the prosecutor's need to advise peace officers as to proper procedure.[45] Moreover, the

---

40. Presently UTAH CODE ANN. § 77-42-1 (1953).

41. 108 Utah 63, 93, 157 P.2d 258, 271 (1945).

42. See notes 15-16 *supra*.

43. Note, *Governmental Appeals of "Dismissals" In Criminal Cases*, 87 HARV. L. REV. 1822, 1838 (1974).

44. Busch & Becker, *supra* note 7, at 12-13.

45. If lower court rulings restricting police conduct cannot be appealed and if inconsistent lower court rulings can be resolved only on an appeal by a defendant, it is most difficult to formulate law enforcement policies. Although it may be argued that erroneous rulings by trial courts will eventually lose their effect as appellate courts consider search and seizure and confession questions raised by defendants, this is an unsatisfactory remedy. When the prosecution is not permitted to appeal, law enforcement officials faced with a restrictive ruling which they feel is erroneous have two choices: They may follow the lower courts decision and abandon the practice in which case an authoritative decision by an appellate court may never be obtained, or they

prosecutor's interest in the development of a uniform body of criminal law may coincide with defense interests because a present state appeal can relieve a future criminal defendant of the burden and expense of appealing in a later case.[46]

Finally, state appeals further the state's interest in regulating the actions of individual trial court judges. A state appeal provides "a review and correction of any despotic or arbitrary dismissal of a case by a judge."[47] The availability of a state appeal could also encourage a conscientious judge to rule on difficult legal defenses because the judge would be aware that a decision in favor of the defendant would not be insulated from appellate review and that any mistake could be corrected.[48]

Some interests of the criminal defendant argue against broadening the state's appellate rights. A restriction on the state's appellate rights works to the advantage of the criminal defendant in a case where appeal is not allowed. However, the defendant's right to take advantage of a trial court's mistakes of law to avoid punishment does not appear to be worthy of the law's protection.

A second argument which could be raised in favor of restricting the state's appellate rights is that an unscrupulous prosecutor could take advantage of his broadened right of appeal and undertake unfounded appeals in order to harass innocent defendants. The most serious objection to this argument is that an unscrupulous prosecutor can more easily and more effectively harass an innocent defendant by bringing successive unfounded charges than by prosecuting unfounded appeals. The restriction of the prosecutor's right of appeal does not significantly protect the right of citizens to be free of prosecutorial harassment.

The defendant's constitutional right to a speedy trial argues most persuasively against an expansion of the state's right of appeal.[49] The defendant has a justifiable expectation that he will not be kept in suspense indefinitely as to the resolution of pending criminal charges and that the proceedings against him will be

---

> may continue the practice . . . solely because of the lack of any vehicle for testing it in the appellate courts. The second course puts the police in the undesirable position of deciding which lower court decisions they will accept and which they will not.

Id. at 15-16, *quoting* 1967 PRESIDENT'S COMMISSION ON LAW ENFORCEMENT AND THM ADMINISTRATION OF JUSTICE, TASK FORCE, THE COURTS 48 (1967).

46. Busch & Becker, *supra* note 7, at 13.

47. State v. Davenport, 30 Utah 2d 298, 300, 517 P.2d 544, 546 (1973) (Crockett, J., dissenting).

48. Note, *Double Jeopardy and Government Appeals in Criminal Cases*, 12 COLUM. J.L. & SOC. PROB. 295, 316 (1976).

49. UTAH CONST. art. I, § 12.

brought to a conclusion as swiftly as possible. This interest of the defendant seems completely incompatible with the grant of appellate rights to the state. The delay in resolution of charges against a defendant which results from a state appeal, however, is mitigated by the fact that a judgment in favor of the defendant is not stayed during the pendency of a state appeal,[50] and the defendant would therefore remain free without bail until his appeal was resolved. If the state seeks an interlocutory appeal before judgment, the appellate court has the power to deny hearing or to render rapidly a decision so as not to impair significantly the defendant's right to a speedy trial.[51] Under these circumstances, the defendant's interest in a speedy conclusion of criminal action does not seem significantly abridged by an appeal which protects the state's interest in punishing the guilty, developing a uniform body of law, and regulating the actions of individual lower court judges. A number of commentators have concluded that a proper balancing of the interests indicates that the state's appellate rights should be expanded.[52]

Because the existing restrictions on the state's right to seek appellate review impair legitimate governmental interests that outweigh the corresponding defense interests, these restrictions should be re-examined. Although these restrictions are alleviated to a degree by the availability of the extraordinary writs, this is not a totally satisfactory solution due to their relatively restricted scope of review and because they are not available as a matter of right.[53]

III. Reforming the State's Authority to Seek Appellate Review

A. *The 1977 Proposed Legislation*

The bill passed by the 1977 General Session of the Utah Legislature would have granted the state the right to appeal from all orders dismissing the charge, orders granting a mistrial, and orders suppressing evidence.[54] This proposal does not meet the objections outlined above and creates some unique problems of its own.

First, the bill would expand the number of situations in which the state would be authorized to seek advisory opinions of the supreme court by allowing appeals from all dismissals. Insofar as the

---

50. Utah Code Ann. § 77-39-8 (1953).

51. *See* Utah R. Civ. P. 72(b); Manwill v. Oyler, 11 Utah 2d 433, 361 P.2d 177 (1961). See notes 65-68 *infra* and accompanying text.

52. Busch & Becker, *supra* note 7; 14 Hous. L. Rev. 735 (1977); 45 U. Cin. L. Rev. 680 (1976).

53. See notes 24-25 *supra* and accompanying text.

54. See note 2 *supra*.

statute would allow appeals of dismissals rendered after jeopardy has attached, but before a verdict has been rendered, it would increase the number of moot cases and hypothetical questions.

Second, the bill, although generous to the state, does not include a right of appeal in every case where the prosecution might desire it. Orders granting a new trial[55] or allowing the defendant to withdraw a previously entered plea of guilty[56] would remain unappealable, although they involve important state interests.

Finally, the provision of the bill which allows the state to appeal from orders granting a mistrial would have no practical purpose. A retrial is available to the prosecutor without an appeal if the defendant consents, or waives his objections to the termination of the first trial, or if there is legally sufficient reason for declaring the mistrial.[57] If a judge erroneously declares a mistrial over the defendant's objection, the discharge of the jury operates as an acquittal and the defendant cannot be retried for the same offense.[58] Therefore, the state would have no incentive to demonstrate on appeal that an order granting a mistrial was erroneous because if such a demonstration were successful it would prohibit the state from retrying the defendant. Even if the state would have a reason to appeal mistrial orders, these orders have been generally considered interlocutory and unappealable.[59] The provision of the bill that would grant the state a right of appeal from mistrial orders appears unnecessary.

Both the proposed legislation and the present Utah law focus exclusively on the types of trial court orders which can be appealed and ignore the judgment which might be rendered on appeal. In effect, they concentrate on creating appellate rights without considering whether the state can obtain an appellate remedy, such as a

55. CAL. PENAL CODE § 1238(a)(3) (West 1970) allows the people to appeal new trial orders. *See* Busch & Becker, *supra* note 7, at 37-41. *See also* 6 SETON HALL L. REV. 376 (1975) (discussion of state appeals of new trial orders).

56. UTAH CODE ANN. § 77-24-3 (1953) provides that a court may allow a defendant to withdraw a plea of guilty. The trial court's actions under this statute are subject to review only for an abuse of discretion. *See, e.g.*, State v. Larson, 560 P.2d 335 (Utah 1977); State v. Lee Lim, 79 Utah 68, 7 P.2d 825 (1932).

57. UTAH CODE ANN. § 76-1-403(4)(1977). This procedure would not violate a defendant's constitutional rights. Illinois v. Somerville, 410 U.S. 458 (1973).

58. State v. Whitman, 93 Utah 557, 74 P.2d 696 (1937).

59. State v. Bauer, 16 Or. App. 443, 519 P.2d 96 (1974). The Utah Code of Criminal Procedure does not grant the defendant the right to appeal a mistrial order. UTAH CODE ANN. § 77-39-3 (1953). Case law suggests that the finality of a judgment might be a constitutional prerequisite to the state's right to appeal. "Whatever else may limit the right of appeal by the state in criminal actions it is beyond question that such right is confined to appeals from final judgments. Const. Utah art. 8 § 9." State v. Thompson, 69 Utah 282, 283, 254 P. 147 (1927).

remand for a new trial. This myopic focus has the inevitable effect of granting the state a right of appeal in cases where it cannot obtain an effective judgment on appeal because the double jeopardy clause of the constitution protects the defendant from further proceedings. Any statute which seeks to properly delimit the state's criminal appellate rights must take into account the availability of a remedy on appeal as well as the type of order from which the appeal was taken. Only in this manner can the statute prevent the pursuit of needless advisory opinions without imposing undue restrictions on the state's right of appeal.

B. *State Appellate Rights Defined by Double Jeopardy Limitations—The Federal Model*

Prior to 1970, the federal statute outlining the United States' right of appeal in criminal cases was similar to the present Utah law; it defined government appellate rights solely in terms of the orders which could be appealed.[60] The act has since been amended to allow the government to appeal from all orders "dismissing an indictment or information as to any one or more counts, except that no appeal shall lie where the double jeopardy clause of the Constitution prohibits further prosecution."[61] The United States Supreme Court has construed this statute to mean

> that Congress decided to rely upon the courts to define the constitutional boundaries rather than create a statutory scheme that might be in some respects narrower or broader than the Fifth Amendment would allow. In light of this background, it seems inescapable that the Congress was determined to avoid creating nonconstitutional bars to the Government's right to appeal. The District Court's order in this case is therefore appealable unless the appeal is barred by the Constitution.[62]

The Utah Legislature should carefully consider this approach to defining the state's appellate rights. The federal statute uses the double jeopardy clause to bar government appeals of moot questions,[63] but fully protects the government's right of appeal when

60. Act of June 25, 1948, ch. 645, 62 Stat. 844 (1948) (as amended and codified at 18 U.S.C. § 3731 (1964 & Supp. IV 1965-68)) allowed the United States to appeal orders setting aside an information or indictment, arresting judgment, and suppressing evidence. The statute did limit the government's right of appeal from orders sustaining a motion in bar to cases where the defendant had not been in jeopardy.

61. 18 U.S.C. § 3731 (1970). The statute also grants the government a right of appeal from orders suppressing evidence.

62. United States v. Wilson, 420 U.S. 332, 339 (1975).

63. The federal courts do not have jurisdiction to hear government appeals in criminal

governmental interests are at stake.[64] A similarly worded state statute such as the following model would allow the state to appeal all final judgments in criminal cases except where further prosecution would put the defendant twice in jeopardy.

Appeal by the state
(1) The state may appeal from any final judgment in a criminal action, provided that no appeal may be taken where the defendant cannot be convicted without further proceedings, and these further proceedings are barred by the laws or constitution of Utah.

The legislature may also wish to consider granting the state the right to appeal orders which do not amount to a final judgment, such as orders suppressing evidence, granting a new trial, or allowing the defendant to withdraw a previously entered plea of guilty.[65] A question remains, however, whether the legislature may constitutionally grant a right of appeal from non-final orders.[66] A better approach would be to authorize the state to petition the supreme court for an interlocutory appeal under Rule 72(b) of the Utah Rules of Civil Procedure. These interlocutory appeals are not taken as a matter of right, and the supreme court may, in its discretion, require the appellant to await the final outcome of the litigation.[67] A statutory provision similar to that which follows would allow the state to have orders reviewed before final judgment in appropriate circumstances.

(2) The Supreme Court may, in its discretion, allow the state permission to take an interlocutory appeal in a criminal case.

The two sections of this suggested statute would give the state optimum criminal appellate rights and would not be difficult to apply.[68]

---

cases where the defendant's rights cannot be affected by the judgment on appeal due to the double jeopardy clause because such an appeal does not present a "case or controversy."

64. Consideration must also be given to the non-constitutional bars to reprosecution contained in the Utah Code of Criminal Procedure. *See, e.g.*, UTAH CODE ANN. §§ 77-31-7 to 9 (1953) (order discharging a defendant so that he may testify against his co-defendant bars further prosecution); UTAH CODE ANN. § 77-51-6 (1953) (dismissal for failure to prosecute is a bar in misdemeanor cases).

65. See notes 55-56 *supra* and accompanying text.

66. See note 59 *supra* and accompanying text.

67. Manwill v. Oyler, 11 Utah 2d 433, 361 P.2d 177 (1961).

68. A statute such as suggested here would give the state an equal opportunity to attack a criminal sentence as a defendant now possesses. In Utah, there is no appellate review of sentences imposed within statutory limits, but a sentence imposed that is not authorized by law can be set aside on appeal. State v. Alexander, 15 Utah 2d 14, 386 P.2d 411 (1963). *See*, Note, *Twice in Jeopardy: Prosecutorial Appeals of Sentences*, 63 VA. L. REV. 325 (1977).

The Utah law governing double jeopardy is reasonably clear and simple, and a statute such as the proposed model would therefore have the clarity required of all jurisdictional statutes. Parties would not need to struggle to establish the court's jurisdiction to hear the appeal and could concentrate instead on the merits of the controversy. The test for determining whether jeopardy has attached should not become so mechanical, however, that it would allow a defendant to manipulate the state's appellate rights. For example, a defendant might attempt to make a dismissal unreviewable by delaying his motions to dismiss until after the start of trial. This kind of problem can be handled in two ways. First, the reviewing court could hold that a delay in making a motion to dismiss prevents the attachment of jeopardy. Thus, a court could hear appeals based on untimely dismissals.[69] Second, the trial court can refuse to rule on the motion until after a verdict has been returned and thus allow the state to appeal an adverse order without placing the defendant in double jeopardy.[70] An appeal, however, could not cause an unwitting defendant to waive his objections to a second trial because no appeal could be heard unless a second trial would be proper.[71]

Should the legislature fail to enact a law similar to that proposed above, state prosecutors can approximate the results of such legislation by two courses of action. First, the state should not seek appellate review in cases where the double jeopardy clause prevents the state from obtaining relief on appeal even if such appeals are authorized by law. These appeals waste scarce judicial resources and do not serve governmental interests well.[72] Second, if a prosecu-

69. *See, e.g.*, Lee v. United States, 97 S. Ct. 2141 (1977); United States v. Appawoo, 553 F.2d 1242 (10th Cir. 1976); United States v. Kehoe, 516 F.2d 78 (5th Cir. 1975).

70. *See* Note, *New York's Illusory Barrier to Government Appeals*, 43 BROOKLYN L. REV. 942 (1975).

71. A defendant may waive his defense of double jeopardy. *In re* Maughan, 6 Utah 167, 21 P. 1088 (1889). For example of how a government appeal might cause a defendant to unwittingly allow his defense of double jeopardy to be waived under present law, see UTAH CODE ANN. § 76-2-303(5) (1977). This statute allows the state to appeal from dismissals granted on entrapment grounds. If a defendant has waived his right to a jury trial prior to filing his motion for dismissal, jeopardy would attach during his "pretrial" hearing on the motion because the first witness would have been sworn and the factfinder would have received evidence. The state cannot constitutionally obtain a remand for trial in this situation. *See Double Jeopardy and Government Appeals*, *supra* note 48, at 330-338. However, a competent attorney might overlook the defense because the defendant had never been made to stand trial.

72. Such an appeal does not aid the state in convicting the guilty because the defendant remains functionally acquitted. See note 38 *supra*. These moot appeals do not aid significantly in developing the law because they are not decided in an adversarial context. See note 39 *supra* and accompanying text. The moot appeal is also ineffective as a sanction against

tor encounters an erroneous decision which cannot be appealed, he should seek review by means of an extraordinary writ. These proceedings would vindicate the governmental interests which could have been served by an appeal, i.e., punishment of the guilty, development of a uniform body of criminal law, and regulation of individual lower court judges.

The Utah Supreme Court could aid in this process of developing an appropriate definition of the state's appellate rights in a number of ways. First, the court should refuse to give advisory opinions. If the state should appeal in a case where relief on appeal is barred by the double jeopardy clause, the court should respond with a short *per curiam* statement that the issues presented need not be resolved because their resolution can no longer affect the parties to the action.[73] Second, when the state legitimately seeks review, the court should be receptive. Petitions for extraordinary writs should be allowed to review a broad range of errors, and statutes authorizing appeal should be liberally construed. If the legislature fails to enact an appropriate statute, *Davenport*[74] should be overruled and the state should be allowed to appeal without express statutory authorization.

The primary responsibility for shaping the state's appellate rights, however, remains with the legislature, and an appropriate statute which focuses on double jeopardy protections for defining the state's appellate rights is the best solution to the problems created by the present system.

WADE LIVINGSTON

---

arbitrary lower court rulings because lower court rulings are not affected by the outcome of the appeal. See note 37 *supra* and accompanying text.

73. The court has indicated an unwillingness to consider moot appeals in civil litigation. Fitzpatrick v. Brown, 41 Utah 139, 124 P. 769 (1912).

74. 30 Utah 2d 298, 517 P.2d 544 (1973). See text at note 16 *supra*.

Prevention of Significant Deterioration of Air Quality: The Clean Air Act Amendments of 1977 and Utah's Power Generating Industry

The policy of prevention of significant deterioration of air quality (PSD) is designed to control industrial emissions in areas where air quality is better than that required by the national ambient air quality standards.[1] PSD policy limits industrial development in these areas to those instances when no "significant" increase in pollution levels would occur as a result of proposed industrialization. Implementation of this policy will have a far-reaching impact on the growth of the energy industry in Utah. This Note will briefly summarize the PSD amendments, examine certain policy considerations underlying PSD, and analyze the application of PSD to two specific case examples, the Intermountain Power Project (IPP) and Utah Power & Light's Huntington and Emery power plants. The Note will also present recommendations for a balanced approach to the accomplishment of the goals and objectives of PSD legislation.

I. BACKGROUND

The Clean Air Act Amendments of 1977[2] require states to in-

1. The national ambient air quality standards for sulfur dioxide and particulate matter are set out in 40 C.F.R. §§ 50.4 to .7 (1976). Approximately eighty percent of the nation's air is estimated to be cleaner than the more stringent of the standards set forth in the regulations. *See* Note, *The Clean Air Act and the Concept of Non-degradation: Sierra Club v. Ruckelshaus*, 2 ECOLOGY L.Q. 801, 825 (1972).

Originally, under the Clean Air Act Amendments of 1970, Pub. L. 91-604, 84 Stat. 1676 (1970) (codified at 42 U.S.C. §§ 1857-57(l) (1970 & Supp. I-V 1971-75)), the Environmental Protection Agency did not require state implementation plans to provide for PSD. In Sierra Club v. Ruckelshaus, 344 F. Supp. 253 (D.D.C. 1972) (memorandum opinion), *aff'd* 4 Envir, Rep. Cas. 1815 (D.C. Cir. 1972) (per curiam) (without opinion), *aff'd* 412 U.S. 541 (1973) (4-4 decision; without opinion), the EPA was enjoined from not requiring provisions for PSD. On the basis of the scant guidelines of the district court's injunction order, EPA promulgated regulations for implementing the policy. 37 Fed. Reg. 19,807, 23,836-7 (1972). The EPA regulations were challenged and upheld in Sierra Club v. EPA, 540 F.2d 1114 (D.C. Cir. 1976). The Supreme Court granted certiorari on the PSD question, 97 S. Ct. 1597 (1977), and the action was still pending when the 1977 amendments to the Clean Air Act were passed. This statute enacted explicit statutory PSD provisions, several of which incorporated the previous EPA regulations. *See* Clean Air Act Amendments of 1977, Pub. L. 95-95, § 127(a), 91 Stat. 731 (1977) (codified at 42 U.S.C.A. §§ 7470-7479 (West Supp. Nov., 1977)). The Supreme Court vacated the judgment below, and remanded the case to the court of appeals for a determination of mootness. 98 S. Ct. 40 (1977).

2. Pub. L. No. 95-95, § 127(a), 91 Stat. 731 (1977) (codified at 42 U.S.C.A. §§ 7470-7479) (West Supp. Nov., 1977)).

clude provisions for PSD in their air quality implementation plans.[3] The Amendments also require adoption of a classification system similar to the one provided for in former EPA regulations.[4] The statute classifies national parks and other specified recreation areas as mandatory Class I areas.[5] All other areas with air cleaner than the national ambient air quality standards are initially classified as Class II. According to an established procedure,[6] a state may redesignate as Class I any area except Indian lands and as Class III any area other than mandatory Class I areas, Indian lands, and areas exceeding 10,000 acres in size which are national monuments or other specified national recreation areas.[7]

Before substantial development occurs in any area subject to PSD legislation,[8] "baseline" concentrations of sulfur dioxide and particulate matter are determined.[9] The statute then specifies for each class the maximum allowable increases for each pollutant over baseline levels for given time periods of exposure.[10] The specifications allow limited increases in Class I areas, moderate increases in Class II areas, and substantial increases in Class III areas. Increases in aggregate projected pollution levels may not exceed the statutory

---

3. 42 U.S.C.A. § 7471 (West Supp. Nov., 1977). Also:

> Each State shall have the primary responsibility for assuring air quality within the entire geographic area comprising such State by submitting an implementation plan for such State which will specify the manner in which the national primary and secondary ambient air quality standards will be achieved and maintained within each air quality control region in such State.

Id. § 7401(a) (West Supp. Nov., 1977).

General procedures and requirements for state implementation plans are enumerated at *id.* §§ 7407-7411.

4. *Compare id.* § 7472(3) *with* 40 C.F.R. § 52.21 (1976).

5. 42 U.S.C.A. § 7472(a) (West Supp. Nov., 1977).

6. *Id.* § 7474.

7. *Id.*

8. Only areas cleaner than the national ambient air quality standards are subject to PSD. *See id.* § 7471.

9. Baseline concentrations are defined as "the ambient concentration levels which exist at the time of the first application for a permit." *Id.* § 7479(4).

Class I increments:

Pollutant	Maximum allowable increase (in micrograms per cubic meter)
Particulate matter:	
Annual geometric mean	5
Twenty-four-hour maximum	10
Sulfur dioxide:	
Annual arithmetic mean	2
Twenty-four-hour maximum	5
Three-hour maximum	25

Class II increments:

Pollutant	Maximum allowable increase (in micrograms per cubic meter)
Particulate matter:	
Annual geometric mean	19
Twenty-four-hour maximum	37
Sulfur dioxide:	
Annual arithmetic mean	20
Twenty-four-hour maximum	91
Three-hour maximum	512

Class III increments:

Pollutant	Maximum allowable increase (in micrograms per cubic meter)
Particulate matter:	
Annual geometric mean	37
Twenty-four-hour maximum	75
Sulfur dioxide:	
Annual arithmetic mean	40
Twenty-four-hour maximum	182
Three-hour maximum	700

Id. § 7473.

increases more than once per year.[11] In addition, no "major emitting facility"[12] may be constructed in any PSD area unless the owner or operator meets eight pre-construction requirements. The requirements include: obtaining a permit from the state certifying that the emissions from the facility will not cause sulfur dioxide or particulate pollution to exceed any of the maximum allowable increments in the area, adoption of the best available control technology, and compliance with provisions regarding "air quality values" in any Class I areas which may be affected.[13] An applicant must also agree to conduct any meteorological monitoring "necessary to determine the effect which emissions from any such facility may have" on the air quality of the area.[14]

The states must transmit permit applications for major emitting facilities to the EPA Administrator. The Administrator provides notice of an application to the Federal Land Manager (or other federal official) responsible for any lands within a Class I area which may be affected by emissions of the plant.[15] The federal official, in consultation with the Administrator, determines whether a proposed facility will have an adverse impact on the "air quality-related values (including visibility)"[16] of a Class I area. If the Federal Land Manager[17] satisfactorily demonstrates to the state that emissions from a proposed facility will have an adverse impact on air quality-related values of the area, a permit for construction cannot be issued, even if allowable increments would not be exceeded.[18] Conversely, if an owner demonstrates to the Federal Land Manager, who so certifies, that no adverse impact on air quality values will

11. *Id.* Pollution from some sources, however, need not be considered. A state may promulgate rules providing that pollutants from certain stationary sources which have converted from the use of petroleum or natural gas to the use of coal and pollutants from new sources outside the United States shall not be taken into account in calculating the increases. Construction and other temporary emission-related sources may also be excluded. *Id.* § 7473(c)(1)(A)-(D).

12. *Id.* § 7479(1) (definition).

13. *Id.* § 7475(a).

14. *Id.* § 7475(a)(7). Heavy reliance is placed on the outcome of the monitoring and the projected pollution levels as determined from meteorological modeling. Data derived from these processes are often a deciding factor in determining whether or not a plant can be built. See text accompanying notes 50-53 and 66-68 *infra.*

15. *Id.* § 7475(d).

16. *Id.* § 7475(d)(2)(B).

17. The EPA Administrator, Federal Land Manager or the Governor of an adjacent state may also make an allegation of adverse impact on a Class I area. Such allegation will prevent construction unless the owner affirmatively demonstrates that the facility will not cause or contribute to pollution levels in excess of the Class I increments, *id.* § 7475(d)(2)(C)(i), or unless a variance is granted. See text accompanying notes 19-26 *infra.*

18. *Id.* § 7475(d)(2)(C)(ii).

result from the emissions, the state may issue a permit even though the facility may cause or contribute to pollution level increments in excess of the Class I maximums.[19] In such a situation, the allowable pollution level increments are essentially identical to the Class II increments.[20]

If an owner is unable to obtain a "no adverse impact" certification from the Federal Land Manager, and the State's Governor does not agree with the federal determination, the Governor may grant a variance.[21] This variance allows the sulfur dioxide concentration to exceed the Class I allowable increase for periods of twenty-four hours, not more than eighteen days a year.[22] Concentrations during those days are not allowed to exceed the Class I increments by more than an additional eight percent of the primary standards in areas of low terrain or fifteen percent in areas of high terrain.[23] In order to obtain this variance, the owner or operator must satisfactorily demonstrate to the Governor, after notice and public hearing, that the facility meets all PSD requirements except the short-term sulfur dioxide Class I increments.[24] Additionally, if any federal *mandatory* Class I areas are affected, the owner must also demonstrate that no adverse impact on "air quality-related values (including visibility)" will occur.[25] If, after the public hearing, the Federal Land Manager still does not concur with the Governor's decision, the decision on whether to grant the variance rests with the President.[26]

II. Analysis of the PSD Amendments

In recent years, considerable attention has been given to quan-

19. *Id.* § 7475(d)(2)(C)(iii).

20. *Id.* § 7475(d)(2)(C)(iv). The increments under the variance allowed in this subsection are the Class II increments as stated in note 10 *supra*, except that the three-hour sulfur dioxide maximum is 325 micrograms per cubic meter instead of 512 as in the Class II increments. This variance is referred to as the Class I relief increments. *See* 123 Cong. Rec. S9,479-80 (daily ed. June 10, 1977).

21. The amendment which provides for this variance represents a modified version of the bill introduced by Congressman Breaux. The variance is often referred to as the Breaux Amendment variance. *See* 42 U.S.C.A. § 7475(d)(2)(D) (West Supp. Nov., 1977).

22. *Id.* § 7475(d)(2)(D)(i)-(iii). Studies indicate that the controlling factor under the PSD amendments is the short-term sulfur dioxide increment requirement. This may be due to uneven terrain characteristics or rare meteorological changes in thermal stratification and stability which may occur only a few times per year. See notes 86, 92 *infra* and accompanying text.

23. 42 U.S.C.A. § 7475(d)(2)(D)(iii).

24. See note 22 *supra.*

25. 42 U.S.C.A. § 7475(d)(2)(D)(i) (West Supp. Nov., 1977).

26. *Id.* § 7475(d)(2)(D)(ii). The Act provides that there shall be no judicial review of the President's decision. *Id.* The statute does not provide for a variance in Class II or III areas. See notes 32-33 *infra* and accompanying text.

tifying the impact of human activity on the environment. Projected trends have stimulated great concern and have prompted several legislative and administrative programs designed to counterbalance these trends. One of the purposes of PSD is "to insure that economic growth will occur in a manner consistent with the preservation of existing clean air resources."[27] Economic and environmental effects entered into the calculation of the allowable increments and the selection of criteria for classification and redesignation.[28] Congress viewed the increments in Classes I, II and III as providing a spectrum of opportunities for industrial development.[29] These increments reflect the congressional judgment that certain pollution levels constitute "significant deterioration" from both an economic and an environmental standpoint. Thus PSD is intended to allow reasonable growth, but the specified air quality requirements, which have incorporated both environmental and economic concerns, determine the extent of that growth.[30]

Unfortunately the PSD policy is, perhaps by necessity,[31] based on assumptions of geographical and meteorological uniformity that do not necessarily conform to reality. For example, supporters of the Amendments cite studies which find that under the Class II incre-

---

27. *Id.* § 7470(3).

28. *See* 123 CONG. REC. S9,260-64 (daily ed. June 9, 1977).

29. *Id.*

30. Several compromises were debated in an attempt to provide more flexibility in the bill. *See* 123 CONG. REC. H4,942-51 (daily ed. May 24, 1977), H5,013-52 (daily ed. May 25, 1977), S9,168-97 (daily ed. June 8, 1977), S9,237-77 (daily ed. June 9, 1977), S9,421-33 (daily ed. June 10, 1977), H8,521-25, 8,548-50 (daily ed. Aug. 3, 1977), S13,700-01, S13,708-09, H8,666-67 (daily ed. Aug. 4, 1977).

One compromise enacted into the bill included the Breaux Amendment, 42 U.S.C.A. § 7475(d)(2)(D) (1977). See notes 21-25 *supra* and accompanying text. Another compromise in the bill included the adoption of Class I relief increments as proposed by the Senate bill, which allow pollution levels to reach essentially the Class II increment levels. The increments were reduced, however, in the case of the controlling factor usually involved, the three-hour sulfur dioxide concentration, from 700 micrograms per cubic meter as in the Senate bill to 325 micrograms per cubic meter. See note 20 *supra* and accompanying text.

Finally, the bill included the adoption of Class III as proposed by the House. The controlling increment, however, three-hour sulfur dioxide concentration, remains at 700 micrograms per cubic meter, the same as the Senate had proposed for Class II. The Class II increment for three-hour sulfur dioxide concentrations is now at 512 micrograms per cubic meter, lower than the Senate bill of 700 but higher than the House bill of 325. *See* 123 CONG. REC. H8,549 (daily ed. Aug. 3, 1977).

31. The rigidity apparent in the PSD amendments is partially explained by the congressional attempt to deal on a nationwide basis with complex interactions between economic and environmental problems that transcend state lines. PSD policy purports to standardize industrial performance and environmental accountability and eliminate competition between the states for industrial development at the expense of the environment. Further, PSD policy claims to provide an adequate compromise among equally deserving and competing nationwide interests. *See* 42 U.S.C.A. § 7470 (West Supp. Nov. 1977).

ments a large coal-fired electric generating plant could be built by complying with the relatively liberal New Source Performance Standards,[32] and that larger plants could be tolerated by using better technology. This argument assumes, however, flat terrain and favorable or ideal meteorological conditions which do not necessarily represent real conditions or which may correlate to areas unfavorable for development from other standpoints. Since a variance is never allowed in a Class II area,[33] the Class II increments may substantially limit plant size in areas of uneven or high terrain or of inconsistent weather patterns. Many of the sites so limited or prohibited from development by these rigid air quality requirements may actually be very favorable sites from standpoints of land and water use, and overall environmental impact. Accordingly, in some areas, the increments may prove to be too restrictive, as the case examples which follow help to illustrate. On the other hand, the increments may be overly permissive in other geographic areas with favorable weather conditions and thus allow extensive industrial development at the expense of other considerations.

Because the policy is premised on assumptions of ideal geographical and meteorological conditions, PSD, in effect, becomes a policy of "site forcing."[34] A proposed plant must comply with the PSD ambient requirements or it will not be built at its proposed site. The effect of this is to force development to take place only at sites where the most favorable conditions for air quality effects exist, with little or no regard to other siting factors, such as the preservation of scenic resources, the socioeconomic impacts on surrounding communities, and the preservation of farm and ranch lands. While the PSD enactments may be satisfactory as guidelines, as statutory requirements they do not provide for needed flexibility to consider the wide variety of factors involved in site selection.

Moreover, excessive environmental restrictions on energy development imposed in an era when growth is needed may lead to "pulling out the stops" later to accommodate the unfulfilled need partially created by the restrictions.[35] The environment, as well as

---

32. *See* note 28 *supra*. *See also* ENVIRONMENTAL RESEARCH & TECHNOLOGY, INC., TECHNICAL EVALUATION OF THE NONDETERIORATION PORTIONS OF PROPOSED CLEAN AIR ACT AMENDMENTS 2-1 (Doc. P-1946-1, Feb. 1977) [hereinafter cited as ENVIRONMENTAL RESEARCH & TECHNOLOGY].

New Source Performance Standards are enumerated in 40 C.F.R. §§ 60.40 to .46 (1977).

33. See notes 66-68 *infra* and accompanying text.

34. The site forcing policy of PSD legislation is analogous to the policy of technology forcing: setting standards of performance currently technologically or economically infeasible and requiring industry to develop the appropriate technology to comply with the standards.

35. *See, e.g.*, Trans-Alaska Pipeline Authorization Act, Pub. L. 93-153, tit. II, 87 Stat. 584 (1973). In 1968, a major oil field was discovered on the north slope of Alaska. A consortium

society, would be losers in the long run if existing restrictions were completely overturned in an attempt to avert a severe economic or energy crisis. Currently an aggregate of several environmental restrictions bridles development.[36] One policy behind these restrictions was to encourage the development of more environmentally acceptable energy and process technology. However, much of the anticipated new technology will likely not come to fruition in time to absorb any significant portion of the rising demand for energy.[37] Society may well be approaching a time when the demand for energy will become so compelling that present environmental standards would be significantly lowered in an effort to keep the economy alive.

There are four alternative methods of avoiding such a crisis: (1) more importation of energy, raw materials, and manufactured goods; (2) substantial conservation and cutting back on the standard of living; (3) allowing a more flexible approach to industrial development; or (4) a combination of any of the above. In determining which path to follow, the best approach is one which encourages and allows time for new technological developments, with the least amount of harm to the environment and to the economic stability of the country in the interim. It may be that society is stuck with the "Pony Express" of coal-fired electric generating plants, pyrometallurgical smelters, and other relatively dirty processes until the "telegraph" of new energy and process technology is more fully de-

---

of oil companies applied to the Department of the Interior for rights-of-way for a pipeline to transport oil across Alaska. Suit was brought alleging violations of the Mineral Leasing Act of 1920, 30 U.S.C. § 185 (1970), in connection with the issuance of the permits and under the National Environmental Policy Act of 1969 (NEPA), 42 U.S.C. §§ 4321-4347 (1970). A preliminary injunction was granted in Wilderness Soc'y v. Hickel, 325 F. Supp. 422 (D.D.C. 1970). After a series of appeals over three years, Congress finally enacted the Trans-Alaska Pipeline Authorization Act, *supra*, amending the Mineral Leasing Act and expressly declaring that no further action under NEPA would be required before construction of the pipeline could proceed. *See* Alyeska Pipeline Serv. Co. v. Wilderness Soc'y, 421 U.S. 240, 241-45 (1975).

36. *See, e.g.*, Wilderness Act of 1964, 16 U.S.C. §§ 1131-1136 (1970 & Supps. II-V 1972-75); Wild and Scenic Rivers Act, 16 U.S.C. §§ 1271-1287 (1970 & Supp. V 1975); Soil and Water Resources Conservation Act of 1977, 16 U.S.C.A. §§ 2001-2009 (West Supp. Feb., 1978); Surface Mining Control and Reclamation Act of 1977, 30 U.S.C.A. §§ 1201-1328 (West Supp. Nov., 1977); National Environmental Policy Act, 42 U.S.C.A. §§ 4321-4361 (West 1977); Federal Land Policy and Management Act of 1976, Pub. L. 94-579, 90 Stat. 2743 (1976).

37. *See* H. GORDON & R. MEADOR, PERSPECTIVES ON THE ENERGY CRISIS 1-6 (1977). *See also* TWENTIETH CENTURY FUND TASK FORCE, PROVIDING FOR ENERGY 3-16 (1977).

> [G]reat . . . investments and more time are necessary to establish the feasibility of commercial production of energy using . . . geothermal power, biomass power, tidal power, wind power, wave power, solar power and fusion. . . . [E]ven if the production of energy from these exotic alternative energy sources proves feasible, it might still pose insurmountable economic or environmental problems.

Id. at 87.

veloped, such as solar energy and hydrometallurgy. The policy in the interim should strive to make as smooth a transition as possible with minimum damage to the economy, environment, or standard of living. A prudent reasonable approach giving all of the factors the consideration they merit in context with each other is essential. Compromise and sacrifice in conservation, importation, industrial development and life style will be necessary.[38] Unfortunately, as the following case examples illustrate, the PSD Amendments will not necessarily help society in its efforts to find a reasonable compromise.

III. Case Analyses

A. *Intermountain Power Project*

The Intermountain Power Project (IPP) proposed by twenty-three cities in Utah and five cities in California consists of a 3,000 megawatt coal-fired electric generating plant. The plant would provide employment for 2,000 coal miners and 550 plant personnel, in addition to 2,600 initial construction jobs. In-lieu property, sales, and use taxes would provide state and local governments with additional revenues of $30-$40 million annually.[39] Estimates indicate that operation of the plant would reduce annual purchases of foreign oil by $800 million.[40] The project would supply three-fourths of the

---

38. Cost-effectiveness criteria must be given more weight in the application of environmental standards. In many cases, the environmental standards now in effect provide only marginal benefits at costs great enough to impede production and the development of new resources. We believe that policymakers should weigh benefits to the environment against costs to the economy or to energy-resource development in establishing environmental standards. When incentives and penalties are imposed to enforce standards, they should also be sensitive to costs. In setting environmental standards, policy makers should treat pollution resulting from productive activities that serve essential policy purposes more leniently than pollution resulting from a nonessential use or process.

Id. at 25-26.

Environmental objectives have sometimes been treated as if they were incompatible with energy objectives. The conflicts over the construction of the trans-Alaskan pipeline and the leasing of federally owned petroleum and coal lands are cases in point . . . . *Similarly, Federal pollution standards have sometimes been set at levels that bear little relation to either the costs or the benefits associated with the required reduction in pollution.* These standards frequently ignore the fact that the *pollution problems facing different regions vary considerably.*

Id. at 13-14 (emphasis added).

39. Intermountain Power Project, Intermountain Power Project and the Clean Air Act Amendments of 1977, 1 (July 11, 1977) [hereinafter cited as IPP & the Clean Air Act]. The payments would be in lieu of tax payments since the participating utilities are publicly owned. *Id.*

40. *See* 123 Cong. Rec. H5,050-51 (daily ed. May 25, 1977). IPP estimates that the plant would reduce oil consumption by 35 million barrels per year, equalling the savings of $1 billion

electrical energy needs of the participating Utah cities and over one-third of the needs of the California cities, serving a total population of 3,700,000.[41]

Coal for the project would come from underground mines in the Wasatch Plateau and Emery coal fields. Minimal land disruption would result from extracting coal from these underground mines, consistent with policies concerning strip mining and land reclamation.[42] This coal is of excellent quality, having a low sulfur content and a relatively high Btu content.[43] Best available control technology (BACT) would be utilized, including electrostatic precipitators which would remove 99.5% of the fly ash, and flue gas desulfurization scrubbers which would remove 90% of the sulfur dioxide and 50% of the remaining fine particulate matter (ash) not removed by the precipitators.[44] As compared to many other sources, the low sulfur coal and pollution control utilized in the plant would provide a relatively clean source of energy. Actual pollution emissions per unit of energy output would be considerably lower than that emitted from many other power plants.[45]

An extensive study by IPP involving engineering feasibility,

---

annually. *See* IPP & THE CLEAN AIR ACT, *supra* note 39, at 2. The dependence on petroleum imports in the United States continues to rise. In 1973, imports amounted to 36.1% of domestic demand. Estimates for 1977 indicate that United States imports rose to 46.1% of demand. *See* EN. USERS REP. (BNA) 81:0303 (1977). Total United States energy consumption of all types of fuel during the first half of 1976 was estimated at 34.6 million barrels per day of crude oil equivalent. *Id.* at 81:0301. Therefore, the 35 million barrels per year savings projected by IPP is approximately equal to one day's energy consumption of all types of fuel in the United States. This saving would also equal approximately four times the forecasted daily oil import level for 1977 of 8.324 million barrels per day.

41. For a more detailed description of the IPP project, *see* 123 CONG. REC. S9,246-47 (daily ed. June 9, 1977), H5,050-51 (daily ed. May 25, 1977).

42. *See* Surface Mining Control and Reclamation Act of 1977, 30 U.S.C.A. §§ 1201-1328 (West Supp. Nov., 1977).

43. The coal will have a 0.55% sulfur content and a heating value of 11,500 Btu/lb. 123 CONG. REC. S9,272-3 (daily ed. June 9, 1977). Typical eastern bituminous coal has a sulfur content of 5-7% and a heating value of 12,000 Btu/lb. *See* Kaplan & Maxwell, *Removal of SO2 From Industrial Waste Gases*, 84 CHEMICAL ENGINEERING 127, 129-30 (October 17, 1977).

44. Total particulate removal would be 99.75% as a result of the precipitators and scrubbers. *See* 123 CONG. REC. S9,273 (daily ed. June 9, 1977).

45. A typical 1000-megawatt power plant burning average coal in the United States emits 139,000 metric tons of sulfur dioxide per year with no flue gas disulfurization, 4,500 metric tons of particulates per year with 97.5% fly ash removal, and 21,000 metric tons of nitrogen oxides. *See* L. HODGES, ENVIRONMENTAL POLLUTION 382 (1977). Assuming that the "typical" plant had a 3,000-megawatt capacity and 90% sulfur dioxide removal, the omissions would be 45,870 tons of sulfur dioxide per year, 14,850 tons of particulates per year, and 69,300 tons of nitrogen oxides per year.

The IPP plant emissions are estimated as follows: 8,081 tons per year of sulfur dioxide; 1,183 tons per year of particulates; and 68,525 tons per year of nitrogen oxides. *See* 123 CONG. REC. S9,273 (daily ed. June 9, 1977) (attachment A to letter from Douglas M. Costle to Sen. Muskie.)

economics, coal and water availability, and environmental effects produced five alternative sites for the proposed plant.[46] A site in Salt Wash, northeast of Capitol Reef National Park, provided the best alternative.[47] The Salt Wash site is in an arid, undeveloped region approximately eight miles northeast of Capitol Reef National Park and sixteen miles northwest of Hanksville in Wayne County, Utah. The North Caineville Mesa surrounds the site and isolates it from nearby communities and the highway. From a standpoint of physical positioning, the site would be highly acceptable as it would be well hidden from most visitors to the area. Construction of railroads to haul coal from the Emery coal fields would cause low adverse impact on the region and low visual intrusion. Transmission lines also would be favorably situated.[48]

The Fremont River and an underground aquifer in the Navajo sandstone in the vicinity would provide 50,000 acre-feet of water per year for the operation of the plant. A reservoir would be constructed to receive water diverted from the river and pumped from the aquifer. The reservoir would provide needed water for local agricultural, industrial, and municipal uses as well as for the plant itself.[49]

Extensive meteorological studies revealed further benefits of the Salt Wash site. Although the site is only eight miles east of the northern end of Capitol Reef National Park, emissions from the plant would seldom effect the park. Prevailing winds, blowing to the northeast, would substantially disperse the plant emissions and carry them toward the San Rafael Swell and away from Capitol Reef.[50] The emissions would have a minimal effect on the San Rafael Swell in a dispersed state. Moreover, results from a study conducted by an independent company for the Bureau of Land Management (BLM) indicated that emissions from the plant at the proposed site would not exceed any Class II increments in any area. Nor would

46. IPP & THE CLEAN AIR ACT, *supra* note 39, at 2.

47. Several significant yet subtle environmental factors contributed to the selection of the Salt Wash site. Primary factors included (1) low potential for air pollution; (2) adequate water availability without disruption of present water consumption patterns; (3) low aesthetic intrusion; and (4) capability to accomodate plant and mining work forces. Secondary factors included minimal disruption of land and aquatic ecology and land use patterns, low potential for archaeological and paleontological finds, and minimal environmental and social impacts in connection with transmission line placement and coal transport. *Id.*

48. *Id.* at 6. For a more detailed description of the Salt Wash site, *see* J. BOWERS, H. CRAMER, & A. ANDERSON, ASSESSMENT OF THE AIR QUALITY IMPACT OF EMISSIONS FROM THE PROPOSED IPP POWER PLANT AT THE PRIMARY AND THREE ALTERNATE SITES 4-8 (TR-77-311-OT, July, 1977) (prepared for the U.S. Dept. of Interior, Bureau of Land Management) (hereinafter cited as BOWERS & CRAMER).

49. Intermountain Power Project (promotion pamphlet, 1976).

50. BOWERS & CRAMER, *supra* note 48, at 18-23.

emissions exceed the Class I annual increments for sulfur dioxide and particulates or the twenty-four-hour maximum increment for particulates in Capitol Reef or any of the other nearby Class I areas.[51] However, due to short-term ground-based inversions and transitions in wind and thermal stratification, studies show that maximum concentrations of sulfur dioxide may exceed the twenty-four and three-hour Class I increments in Capitol Reef during a limited number of days per year.[52] Emissions from a plant situated at Salt Wash would not exceed any of the Class I increments at any other park or recreation area.[53]

Assuming that the effects on vegetation, wildlife, and other ecological concerns are already adequately taken into consideration in establishing the increments for Class I and Class II, the only values left to consider for a mandatory Class I area are visibility and aesthetic concerns. Under clear air conditions, visibility at Capitol Reef extends a distance of eighty-seven miles. Studies indicate, however, that naturally occurring haze and windblown dust reduce visibility in the area to less than forty miles approximately twenty percent of the year and to less than twenty miles approximately twelve percent of the year. The studies estimate that the Salt Wash plant would reduce the clear air distance in the park to eighty-two miles two to three percent of the time.[54] These temporary visibility decreases would usually occur in winter or during early morning or late evening hours when moisture is present in the atmosphere, times corresponding to typically low tourism activity.[55]

Considering a balance of several of the environmental, economic, and energy data available, construction of the proposed plant at the Salt Wash site appears to conform to the policies of energy independence, coal utilization,[56] low land disruption, as well as air quality conservation. Unless "significant deterioration" of air quality encompasses a conjectural few days per year during which

51. *Id.* at iv, 39-54.

52. *Id.* at 43-54. Other indications are that the emissions from the plant could exceed the Class I sulfur dioxide increments in Capitol Reef for periods of twenty-four hours or less on approximately thirteen days per year. *See* 123 CONG. REC. S9,246 (dailey ed. June 9, 1977) (remarks of Sen. Hatch).

53. BOWERS & CRAMER, *supra* note 48, at 50-53.

54. The distance across the park is only sixty-six miles. IPP AND THE CLEAN AIR ACT, Questions and Answers 3. *See also* 123 CONG. REC. S9,246-47 (daily ed. June 9, 1977) (remarks of Sen. Hatch).

55. 123 CONG. REC. S9,249-51 (daily ed. June 9, 1977).

56. Part of the Carter administration's plan to reduce dependence on foreign sources is to increase domestic coal production from the current rate of 665 million tons per year to 1.2 billion tons per year by 1985, and to triple the current rate by the end of this century. 123 CONG. REC. S9,260 (daily ed. June 9, 1977) (remarks of Sen. Johnston).

sulfur dioxide concentrations in a national park *may* reach one/twenty-fifth to one-tenth of the national ambient air quality standards,[57] construction of the plant would also be in conformance with PSD policy.[58]

Nevertheless, the Secretary of the Interior[59] has refused to allow siting of the IPP plant at Salt Wash. Under the 1977 Clean Air Act Amendments, the Secretary could have granted a variance for the relatively minor violations of the Class I standards involved in the site but has refused to do so. As a result, the IPP plant will probably be placed near Lynndyl, Utah, a site chosen by the State Task Force.[60] It is significant that the Secretary chose not to use the flexibility provided for in the Act to approve the Salt Wash location. Given that the Salt Wash site at least partially motivated the development of the Act's variance provisions,[61] the Secretary's refusal to grant a variance for the site may indicate that even the small amount of flexibility built into the PSD policy is more one of form than of substance.[62]

---

57. The twenty-four-hour increment in Capitol Reef is estimated to reach 14 micrograms per cubic meter on a few days per year when the allowable Class I increment is five and the primary ambient standard is 365. The three-hour increment is expected to reach 124 micrograms per cubic meter when the allowable increment is 25 and the ambient standard 1300. *See* BOWERS & CRAMER, *supra* note 48, at 53.

58. A fourth alternative site adjacent to a major highway between Price and Green River was also studied. Models indicate that emissions from a plant erected at that site would not exceed any Class I increment in any of the presently designated Class I areas. J. BOWERS, H. CRAMER, & A. ANDERSON, ASSESSMENT OF THE AIR QUALITY IMPACT OF EMISSIONS FOR THE PROPOSED IPP POWER PLANT AT AN ALTERNATE SITE NORTHWEST OF GREEN RIVER, UTAH ii-vi (TR-77-311-02, August, 1977) (prepared for the U.S. Dept. of the Interior, Bureau of Land Management). This site, however, could present several difficulties in engineering and economics as well as distinct adverse environmental effects. Specifically, increased environmental impacts would result from longer coal haul roads, transmission lines, and cooling water supply lines. The plant would be visible from the highway and would potentially detract from the natural state of the land as viewed by tourists passing through the area. IPP AND THE CLEAN AIR ACT, Questions and Answers 6. Other factors, however, such as the capacity of the site to absorb growth from the influx of people which would occur attendant to the IPP project have not been specifically addressed. While admittedly they too are factors worthy of careful consideration in plant siting, they are beyond the immediate scope of this Note.

59. The Secretary of the Interior is the Federal Land Manager over the land in question.

60. *See* Deseret News, Feb. 28, 1978, § B, at 1, 2.

61. The IPP plant siting problems are intimately related to the passage of the Breaux Amendment. Congress had on the floor the Clean Air Act bill which required strict compliance with its specificiations. A group seeking to amend the statute presented to Congress one specific fact situation, the IPP plant at Salt Wash which, under the bill, would not be allowed. The benefits of the plant at that site were weighed against the detriments in lengthy and detailed debates. *See* 123 CONG. REC. H5,013-52 (daily ed. May 25, 1977); *Id.* S9,237-77 (daily ed. June 9, 1977). As a result of the debate, the Breaux Amendment to the bill, which would allow plant construction at Salt Wash with Federal Land Manager approval, was passed. In other words, Congress approved the Salt Wash site subject to Federal Land Manager discretion. See notes 21-26 *supra* and accompanying text.

62. Even though the variance provided by the Breaux Amendment would technically

B. Utah Power and Light

In Huntington Canyon, twenty-nine miles southwest of Price, Utah, Utah Power and Light began contruction of the first unit of its Huntington Power Plant before EPA promulgated PSD regulations classifying the land as Class II.[63] Huntington Canyon, a small box canyon, lies on the edge of the Wasatch Plateau. The plant is located on a bench at the mouth of the canyon near the confluence with Deer Creek Canyon. The physical positioning of the plant site is ideal for minimal visual intrusion since it is substantially hidden inside the box canyon. Although the tops of the stacks can be seen from the nearby town of Huntington, the plant causes no other significant visual impact outside the canyon mouth.

Coal for the plant moves over a narrow covered conveyor belt resembling a pipeline which transverses a ridge to the southwest. The conveyor belt carries coal from Peabody's Deer Creek Mine, an underground coal mine situated in Deer Creek Canyon two and one-half miles from the plant. The coal is quite clean and is mined entirely underground, consistent with conservation policy regarding land preservation.[64] Water comes from a nearby stream and small reservoir. Subsequent to the promulgation of the EPA PSD regulations, Utah Power and Light added a second unit identical to the first.[65]

Utah Power and Light initially planned for development of nearby coal and water supplies and disposal facilities to accommodate four generating units at the site with a total capacity of 2,000 megawatts. However, meteorological modeling performed for the

---

have allowed the siting of the IPP plant at Salt Wash based on data available at the time of enactment, revisions of the Bowers and Cramer studies, *supra* notes 48-55, indicate that the plant may be in non-compliance at Capitol Reef even with the variance. The final report, J. BOWERS, H. CRAMER, & A. ANDERSON, ASSESSMENT OF THE AIR QUALITY IMPACT OF EMISSIONS FROM THE PROPOSED IPP POWER PLANT AT THE PRIMARY AND SIX ALTERNATE SITES (TR-78-311-01, January, 1978) (prepared for the U.S. Dept. of Interior, Bureau of Land Management) [hereinafter cited as BOWERS & CRAMER FINAL], indicates that the three-hour sulfur dioxide concentration would exceed the Class I increment at Capitol Reef thirty-four days per year, rather than only thirteen as anticipated by the supporters of the Breaux Amendment. The twenty-four-hour maximum in Capitol Reef would exceed the Class I increment only twelve days per year. *Id.* at 112. These conclusions are refuted by IPP. *See* Deseret News, note 60 *supra*.

63. Construction began March 8, 1971. *See* BUREAU OF LAND MANAGEMENT, ENVIRONMENTAL STATEMENT: SECOND UNIT HUNTINGTON CANYON GENERATING STATION AND 345 KV TRANSMISSION LINE (1975) [hereinafter cited as HUNTINGTON ENVIRONMENTAL STATEMENT]. The EPA regulations came out somewhat later. See note 1 *supra*.

64. See note 42 *supra*.

65. HUNTINGTON ENVIRONMENTAL STATEMENT, *supra* note 63, at A-1 to 3. The first unit built had no scrubber, while the second unit planned will utilize a scrubber to remove eighty percent of the sulfur dioxide. *Id.* See notes 66-68 *infra*.

canyon and surrounding high terrain areas projected that there would be violations of the sulfur dioxide Class II increments in the canyon and on the nearby cliffs.[66] If the high terrain did not surround the plant, the pollutants would be widely dispersed and the violations might not occur. Yet it is the cliffs that protect the scenic value of the area by hiding the plant. The cliffs are made up of impervious rock with little vegetation and there is little potential for any decomposition from the acidic sulfur dioxide emissions. Yet because the plumes impinge upon them occasionally, the Class II increment may be exceeded.[67] Consequently, an extremely advantageous site for further development consistent with reasonable and wise environmental planning may be prohibited from consideration due to the rigid air quality requirements and the reliance on meterological modeling to determine site suitability.[68]

---

66. The State Air Quality Department conducted a modeling study of Huntington Canyon using the EPA "Valley" model and 1975 data. The study used sulfur dioxide emission rates of two units, one with a scrubber (94.5 grams per second) and one without a scrubber (407.4 grams per second). The maximum annual sulfur dioxide concentration predicted to occur was 42.0 micrograms per cubic meter (the Class II allowable increment is 20). The maximum twenty-four-hour average concentration predicted was 886 micrograms per cubic meter (the Class II allowable increment is 91). A projection of a three-hour maximum was not made. *See* Huntington Modeling Data—1975-1977 (unpublished record at the Utah State Air Quality Department).

Another study, using almost equivalent emission rates for the two units indicated that the maximum annual sulfur dioxide concentration would be 14.1 micrograms per cubic meter (below the Class II increment of 20). The maximum twenty-four-hour average would be 305 micrograms per cubic meter (above the Class II increment of 91) while the maximum three-hour concentration would be 1578 (above the Class II increment of 512 and even above the national ambient standard of 1300). The high concentrations in his study correspond to the high terrain and the short-term exposure periods. *See* H. CRAMER & J. BOWERS, ASSESSMENT OF THE AIR QUALITY IMPACT OF EMISSIONS FROM THE EMERY AND HUNTINGTON POWER PLANTS 18, 48-54 (TR-76-114-01, August, 1976) (prepared for the U.S. Dept. of Interior, Bureau of Land Management) [hereinafter cited as EMERY AND HUNTINGTON STUDY].

67. *Id.*

68. Notwithstanding the predictions of high sulfur dioxide concentrations and the subsequent reliance upon them in site evaluation, actual operation of a plant often produces a contrary outcome. Observations made in conjunction with the studies described in note 66 *supra*, and the on-going monitoring of the Huntington plant emissions, measured considerably lower concentrations than the projected maximum concentrations. In the EMERY AND HUNTINGTON STUDY, *supra* note 66, at a location on the canyon floor, the twenty-four and three-hour maximum observed sulfur dioxide concentrations were 69 and 251 micrograms per cubic meter, respectively (lower than the Class II increments of 91 and 512). In two locations on the ridges surrounding the canyon, though the measured twenty-four-hour maximum concentrations of 216 and 161 micrograms per cubic meter were in excess of the Class II increment of 91, the measured three-hour maximum concentrations of 472 and 502 were below the Class II increment of 512.

The State Air Quality Department monitoring during 1975 (when only one unit without a scrubber, emitting sulfur dioxide at eighty percent of the rate considered in the modeling was operating) showed quite low sulfur dioxide concentrations. The annual arithmetic mean measured on the canyon floor was less than .0005 parts per million (below approximately 1.3

The additional capacity initially planned for Huntington will now come from the Emery plant currently under consideration. Utah Power and Light made the Emery site choice primarily with the EPA PSD regulations in mind.[69] The plant is located about fourteen miles southeast of the Huntington plant in the broad and relatively flat Castle Valley, out in the open and removed from elevated terrain.[70] It is visible to visitors and residents, and is close to towns.[71] Coal haul roads and cooling water supply lines will have to be constructed to service the plant, causing increased costs over the existing Huntington facility.[72] Nine coal haul trucks per hour, sixteen hours per day, will pass along the coal haul road during operation, further increasing costs and energy consumption.[73] Consequently, the plant as well as its associated activity may disrupt the ecology of the area considerably more than the Huntington plant site,[74] even though it probably will comply with the Class II increments.[75]

The arbitrary "site forcing" reflected in the choice of the Emery plant site and imposed by the PSD Amendments illustrates the failure to incorporate cogent industrial and environmental concerns. The Huntington site, selected by industry with many economic and environmental factors in mind, is superior in many ways to the Emery site. The Emery site, chosen for its potential compliance with the PSD increments, has relatively few other notable virtues.

IV. Analysis of PSD as Applied, and Recommendations

As mandated by statute,[76] meteorological modeling is used to

---

micrograms per cubic meter) which is even below the Class I increment of two micrograms per cubic meter. The twenty-four-hour maximum observed concentration was .01 parts per million (approximately 26 micrograms per cubic meter), well below the Class II increment of 91. The three-hour observed maximum of .02 parts per million (approximately 52 micrograms per cubic meter) was also well below the Class II increment of 512. Other measurements in 1975 on the ridge surrounding the plant and at other sites on the valley floor reached as high as .03 parts per million (80 micrograms per cubic meter) for a twenty-four-hour period, still below the Class II increment. Subsequent data has not been representative due to plant shutdowns in 1976 and 1977. *See* Huntington Modeling Data—1975-1977, *supra* note 66.

69. Bureau of Land Management, U.S. Dep't of Interior, Final Environmental Statement—Emery Appendix I-2 to I-5 (1977).

70. *Id.* at 3-59 to 64.

71. *Id.* at 1-24 to 26.

72. *Id.* at 3-94.

73. *Id.* at 3-90 to 94.

74. Land occupied by the plant and the roads and lines will be removed from agricultural production. Loss of grazing land and a reduction of beef production equivalent to an annual consumption for 3,280 persons would occur. Agricultural production loss would equal a loss of beef production for 2,720 persons. *Id.* at 3-70.

75. *Id.* at 5-2. *See also* Emery and Huntington Study, *supra* note 66, at 36.

76. 42 U.S.C.A. § 7475(a), (e) (West Supp. Nov., 1977). The statute requires monitoring

evaluate the air quality impacts of proposed developments and determine their compliance with PSD legislation. Controversy surrounding the accuracy and reliability of state-of-the-art meteorological modeling provides grounds for significant confrontation between industry and environmentalists. Each side tends to criticize modeling as giving results which favor the other. Realistically, modeling merely presents a sophisticated estimate of what *may* happen. Data derived from the modeling are, especially in areas of diverse terrain and weather conditions, tenuous.[77] Reasonableness would seem to require that before a final decision on a plant site is made, theoretical data derived from modeling should be considered in light of more well known concrete information, such as land use effects, coal and water availability, and the ecological effects of their utilization at the proposed site. Arguably, the variances allowed by the Clean Air Act provide the flexibility necessary to include this information in the decision-making process.

However, general fear that meteorological models underestimate air quality effects and an awareness of the irreplaceable nature of national parks will impose pressure on Federal Land Managers and Governors not to exercise their discretion under the Act to grant Class I variances when a national park or other areas of high aesthetic value would be affected.[78] It is probable that the variance provisions of the Act will only be used to allow construction of nonconforming plants where slight and conjectural adverse air quality impacts are overwhelmingly counterbalanced by energy needs, economics, and environmental benefits other than air quality.

The latent inequities and undue rigidity of the PSD enactments become apparent as their potential application is examined. States with large proportions of federal mandatory Class I areas, potential discretionary Class I areas, or large areas of potential "wilderness" regions may be burdened with limited opportunities for industrial development.[79] States with meteorological irregularities or diverse

---

to be done such as is necessary to determine the effect of proposed plant emissions, as well as conducting meteorological modeling studies. *Id.*

77. For a critique of certain modeling techniques and an explanation of their limitations in regard to diverse terrain and weather patterns, *see* ENVIRONMENTAL RESEARCH & TECHNOLOGY, *supra* note 32, at 5-1 to 3.

78. If the Secretary of the Interior does choose to exercise his discretion to grant a Class I variance for a national park, he may be subject to suit for violation of his duty under the National Parks Service Act. Congress has mandated the preservation of the historic, scenic, and natural objects and wildlife within the national parks for "the enjoyment of future generations." *See* 16 U.S.C. § 1 (1970). Arguably, allowing any increased pollution in a national park violates this mandate.

79. Estimates indicate that buffer zones for Class I areas will be 25-150 miles, depending on terrain and modeling assumptions. *See* ENVIRONMENTAL RESEARCH & TECHNOLOGY, THE

terrain may encounter similar limitations even in Class II areas.[80] Intricate and difficult calculations involving interactions between economic variables and imposed environmental parameters other than air standards already considerably restrict industrial siting.[81] When rigid air quality standards are compounded with existing restrictions, vast regions of several states may be precluded from any development. Consequently, there may be many Class II areas, particularly high terrain areas, which for several reasons (other than strict compliance with Class II increments) would be desirable and advantageous for development, but which will be excluded from consideration.

The negative impact of such exclusion can be demonstrated by examination of the power generating industry in Utah. Utah Power and Light currently purchases approximately twenty percent of its power from outside sources such as the Northwest Power Pool, which provides power from hydroelectric facilities. The growth rate for Utah Power and Light's service area is approximately eight percent per year.[82] Under the PSD enactments, with the accompanying magnitude of mandatory and potential Class I areas in Utah and attendant difficulties in siting near such an area, Utah Power and Light estimates that there are only two regions left within the state for power plant siting.[83] These regions comprise only seven general sites in which reasonably scaled power plants could be constructed, unless substantial areas were redesignated as Class III.[84] All of these sites would be utilized by Utah Power and Light alone within the next thirty years in order to provide power for its service area at its current rate of growth. This leaves little or no development open to

---

IMPACT OF SIGNIFICANT DETERIORATION PROPOSALS ON THE SITING OF ELECTRIC GENERATING FACILITIES—DOCUMENTATION OF ANALYSES UNDERTAKEN BETWEEN JULY 1975 AND SEPTEMBER 1976, 6 (Doc. No. P-1946-2, February, 1977).

80. The combination of many nearby Class I areas and diverse terrain will impose extreme limitations on new development in many areas. Plume impingement on higher elevations would remove several areas from consideration. See the IPP and Utah Power and Light case examples at notes 39-75 *supra* and accompanying text.

81. Economics, engineering feasibility, resource availability, land use considerations and many other factors eliminate many areas from consideration for development. See note 84 *infra* and accompanying text.

82. Presentation by Utah Power and Light Spokesman, Juab County Commissioners' meeting, Nephi, Utah (Nov. 11, 1977). A growth rate of such a magnitude means that demand will double every ten years.

83. *Id.*

84. *Id.* The areas considered are within the Price-Emery-Green River corridor, in the Delta-Lynndyl region and hills west of Nephi. An area southeast of Vernal is also feasible but was not considered in order to allow for speculative oil shale development which will need to use the PSD increments. Mountain peaks and high plateaus were excluded from consideration due to engineering problems and other practical aspects. *Id.*

other companies.[85] If power plant development in Utah to supply other areas is desirable due to availability of low sulfur coal and other favorable environmental and economic factors, a reasonable approach would encourage that development as the need arises. Such development would unfetter hydroelectric power for other regions. If Utah's own demands consume all of the allowable increments in all of the available areas, no power could be generated for other use.

Redesignation of certain areas to Class III presents one solution to this dilemma. Redesignation to Class III would allow pollution to reach half of the levels of the national ambient air quality standards, and would probably attract heavy industrialization to the redesignated area. This may or may not be desirable. If the area were not particularly adapted for such development, or if redesignation became a frequent procedure for accommodating development, it would certainly be inconsistent with the Clean Air Act policy "to protect and enhance" the nation's air quality.

Alternatives to redesignation should be devised for allowing more flexibility in the application of the PSD policy. Casting the PSD increments as prima facie standards, or allowing a Class II variance to permit local air quality to exceed the short-term sulfur dioxide increments a small percentage of the time might prove to both protect the environment and allow for economic growth.[86] If the PSD requirements were a prima facie standard instead of a statutory mandate, other beneficial environmental factors could be brought forth and weighed against air quality effects. Proper emphasis could attach to each factor in light of the economic and environmental considerations involved. Implementation of a Class II variance could achieve the same result. Further, a Class II variance would not result in additional pollutants being emitted into the atmosphere. A plant under a Class II high terrain variance in an area of uneven topography or variable meteorological conditions would emit no more than a plant in a flat Class II area when each is required to employ "best available control technology,"[87] and burns the same quality of coal.

---

85. *Id.*

86. Modeling indicates that:

> [F]or a small percentage of the 3 and 24 hour periods in each year, local compliance with the short-term increment limits will be strongly related to the details of topography and meteorology in each specific region of the country. . . . [T]he question of site suitability . . . will be controlled by details of local topography, rather than by general requirements for strict emission limits and for limits to total growth in each region.

ENVIRONMENTAL RESEARCH & TECHNOLOGY, *supra* note 32, at 2-5.

87. *See* 42 U.S.C.A. § 7475(a)(4) (West Supp. Nov., 1977).

Alternatively, a selective approach to redesignation to Class III could be applied in unique areas.[88] Under this aproach, a small area, such as a high mountain slope or cliff, affected by occasional impingements of plumes could be redesignated as Class III. The redesignation would allow the immediate air-ground interface to exceed the Class II increment but would not allow any further industrialization. The redesignation would involve no adverse impacts on any other area, and would merely serve as a normalizing factor for states with diverse terrain. The redesignation so suggested would have a result similar to the Class II variance proposal.

Reducing PSD to a prima facie standard, permitting a Class II variance similar to the Class I variance, or a specialized approach for redesignation to Class III under certain conditions would yield several advantages consistent with conservationist PSD policy. Primarily, more land would be amenable to well planned development in states with high terrain without encouraging heavy industrialization. At the same time, a Class II variance would relieve pressure to use areas encircling national parks and other similar Class I areas. A Class II variance would further conform to the "protect and enhance" policy, since it would allow a reduction of pollution burdens in dirty areas by providing well chosen areas in which development could occur. These development areas could absorb on-going production or industry expansion, enabling the "enhancement" of non-attainment areas,[89] where the bulk of the population lives, to begin to approach the standards set for health and welfare.[90] In addition, existing hydroelectric power operating units in other states would become available to absorb growth in their service areas.

The siting advantages of a Class II variance would be striking. A general policy of air quality preservation coupled with other environmental concerns would control site selection, rather than speculation among meteorologists as to which models best demonstrate the proper "worst case" analysis.[91] Constructive enviornmental planning could result, with appropriate emphasis on land use effects, ecological impacts, and air and water quality control. Without

88. *See* 123 CONG. REC. H8,666, S13,708 (daily ed. Aug. 4, 1977).

89. A non-attainment area is one in which pollution exceeds the national ambient air quality standards. *See* 42 U.S.C.A. §§ 7501-7508 (West Supp. Nov., 1977).

90. Non-attainment of the National Ambient Air Quality Standards in "dirty air" areas continues to present considerable difficulty in clean-up. *See* 123 CONG. REC. H4,951 (daily ed. May 24, 1977).

91. A "worst case" analysis comprises the assumptions made to determine which models reflect most accurately the expected effect of unfavorable meteorological conditions. *See e.g.*, BOWERS & CRAMER FINAL, *supra* note 62, at 18.

a variance, the use of speculative models could breed lengthy disputes and will inevitably control site suitability in high terrain areas. Heavy emphasis on models alone does not seem to foster responsible environmental planning.[92] However, unless some additional provision is incorporated into the Clean Air Act, the country may be forced to accept this one-parameter environmental planning.

V. CONCLUSION

While the Clean Air Act Amendments of 1977, including the Class I variance, may provide a certain necessary degree of flexibility in the PSD policy, the Act goes only half way. The general policy of "protect and enhance" includes the prevention of significant deterioration of any area, as well as the clean-up of regions with dirty air. This policy calls for essentially no development in and around mandatory Class I areas, wise planning of development in presently "clean" areas, and simultaneous reduction of pollution burdens in industrial centers. Seemingly, some development in "clean" areas will be essential in order to facilitate pollution reduction in developed areas. This development could more easily occur under a Class II variance provision or some other method of achieving flexibility in the Act in addition to the Class I variance.[93]

Economic growth "consistent with the preservation of existing clean air resources" requires extensive planning and efficient utilization of land, water, and mineral resources. Local environmental characteristics often dictate the best location for industry. Case-by-case analyses are needed to determine such locations. While PSD enactments provide a well-planned framework for air quality considerations, some flexibility in the requirements should be allowed in individual circumstances where other considerations may override small adverse air quality impacts. Casting the PSD enactments as a prima facie standard, allowing a Class II variance, or

---

92. [Modeling projections] during a few "worst hours" per year can act as an absolute veto on the environmental suitability of a proposed site regardless of other environmental, economic or social factors. Even if a site were best suited relative to other environmental considerations, it would be prohibited from development. . . . The presence or absence of [very unusual conditions] in a specific meteorological record can control the selection of a site which might otherwise be most suitable in terms of environmental considerations.

ENVIRONMENTAL RESEARCH & TECHNOLOGY, *supra* note 32, at 2-5.

93. Development may likely occur more reasonably and easily even under the present Class I and Class II increments when modeling technology is further developed to conform more adequately to reality in areas of diverse terrain and weather conditions. See text accompanying notes 66-68 *supra*.

granting specialized redesignation to Class III areas could provide the needed flexibility.

The breadth of interaction among economic, energy, and environmental considerations, coupled with the complex impacts of industrial development, require detailed ad hoc problem solving. While environmental concerns deserve strict guidelines, absolute statutory standards for air quality ignore the complexity of the problems, and can violate the policies from which they spring. Other environmental concerns, particularly, deserve careful consideration in PSD administration.

GEORGE S. YOUNG

Symbolic Speech and Compelled Expression: *Wooley v. Maynard*

I. INTRODUCTION

In *Wooley v. Maynard*,[1] the state of New Hampshire charged one of its residents, George Maynard, with covering over the state motto, "Live Free or Die," on his license plates. Maynard, a Jehovah's Witness, found the motto objectionable to his religious and political beliefs.[2] He therefore obscured the motto to avoid participating in its dissemination, thus violating New Hampshire law.[3] A state district court routinely found him guilty for having violated that law.[4] Firm in his convictions, Maynard continued to cover the motto. Accordingly, the New Hampshire police continued to issue citations.[5] Faced with the potential of being arrested everytime he drove his car, Maynard and his wife sought declaratory and injunctive relief in federal district court.[6]

The district court enjoined the state from prosecuting the Maynards after concluding that their conduct in obscuring the state motto was "smybolic speech" protected by the first amendment.[7] On appeal, the United States Supreme Court affirmed, but on the

---

1. 97 S.Ct. 1428 (1977).

2. Maynard described his objections to the state motto as follows:

> [B]y religious training and belief, I believe that my "government"—Jehovah's Kingdom—offers everlasting life. It would be contrary to that belief to give up my life for the state, even if it meant living in bondage. . . .
>
> I also disagree with the motto on political grounds. I believe that life is more precious than freedom. . . . I have read that it was passed by the legislature in 1969 to answer critics of the Vietnam War. . . .

Affidavit of George Maynard, Appendix to Brief for Appellant at 3, Wooley v. Maynard, 97 S. Ct. 1428 (1977) [hereinafter cited as Appendix].

3. N. H. REV.STAT. ANN. § 263:1 (Supp. 1975). "[N]umber plates for non-commercial vehicles shall have the state motto 'live free or die' written thereon."

N. H. REV. STAT. ANN. § 262:27 (Supp. 1975). "Any person . . . who knowingly obscures or permits to be obscured the figures or letters on any number plate attached to any motor vehicle . . . shall be guilty of a misdemeanor."

4. Although sympathetic, the judge felt obligated to find Maynard guilty because the New Hampshire Supreme Court had already held that a person convicted of obscuring the motto on the state's license plates was not deprived of any first amendment rights. State v. Hoskin, 122 N.H. 332, 295 A.2d 454 (1972). See text accompanying notes 36-39 *infra*. *See also* Transcript of Proceedings in the United States District Court for the District of New Hampshire, Civil Action no. 75-57, December 6, 1974, *reprinted in* Brief for Appellees, Appendix A, Wooley v. Maynard, 97 S. Ct. 1428 (1977).

5. The first citation had been issued on November 27, 1974, a second on December 29, 1974, and a third on January 3, 1974. Maynard had served fifteen days as a result of the first two convictions. *See* Appendix, *supra* note 2, at 15, 16, 18, 64. *See also* Wooley v. Maynard, 97 S. Ct. 1428, 1432 (1977).

6. The Maynards' complaint sought declaratory and injunctive relief under 42 U.S.C. 1983 (1970). *See also* Appendix, *supra* note 1, at 5, 6.

7. Maynard v. Wooley, 406 F. Supp. 1381 (1976).

"more appropriate First Amendment grounds" that Maynards' right "to refrain from speaking" had been violated.[8] The Supreme Court held that a state may not "constitutionally require an individual to participate in the dissemination of an ideological message by displaying it on his private property in a manner and for the express purpose that it be observed and read by the public."[9]

This Comment will show that the grounds relied upon by the Supreme Court—freedom from compelled expression—are "more appropriate" only because of the narrow judicial definition of "symbolic speech" which has emerged in recent years. Indeed, compelled expression grounds need *not* be "more appropriate" than symbolic speech grounds when the underlying first amendment interest—freeedom of expression—is considered. By combining methods previously used to test for one ground or the other, this Comment develops a single test which may be used to identify when this underlying first amendment interest exists. Once identified, another test may be used to determine if a state's countervailing interest is justified in overriding the individual's interest.

II. BACKGROUND

A. *Constitutional Law*

The first amendment protects freedom of expression by allowing expression of one's own ideas and by freeing one from compelled expression of another's ideas.[10] The ideas, or thoughts, are protected absolutely.[11] Their expression, however, may be limited depending upon the form of expression employed and the nature of the government's interest in restricting or forcing that expression.[12]

---

8. Wooley v. Maynard, 97 S. Ct. 1428, 1434-35 (1977).

9. *Id.*

10. *See, e.g.*, EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION (1970). For an excellent judicial defense of the right of free expression, see Whitney v. California, 274 U.S. 357, 372 (1927) (Brandeis, J., concurring). *See also* West Virginia State Bd. of Educ. v. Barnette, 319 U.S. 624 (1943).

11. United States v. Ballard, 322 U.S. 78, 86 (1944); Cantwell v. Connecticut, 310 U.S. 296 (1940); Palko v. Connecticut, 302 U.S. 319 (1937). *See also* EMERSON, *supra* note 10, at 9.

12. *See, e.g.*, Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (speech may be restricted where it is likely to produce or incite imminent unlawful action); Communist Party v. Subversive Activities Control Bd., 367 U.S. 1 (1961) (right of free speech outweighed by competing governmental interest); Roth v. United States, 354 U.S. 476 (1957) (illegal acts, such as the dissemination of obscene material are not protected by the first amendment); Schenck v. United States, 249 U.S. 47, 52 (1919) (expression which constitutes a clear and present danger may be limited). The government's interest which must appear before any control of expression is justified has been characterized by many descriptive terms, such as: "compelling," "substantial," "subordinating," "paramount," "cogent," and "strong." *See* United States v. O'Brien, 391 U.S. 367, 376 (1968) (citations omitted).

One form of expression recognized by the Court in recent years is voluntary expression through non-verbal actions, or "symbolic speech."[13] Symbolic speech exists whenever: (1) there is non-verbal conduct intended to convey a particularized message, and (2) there is a strong likelihood that the message will be understood.[14] Conduct such as displaying a red flag,[15] burning a draft card,[16] wearing a black arm band,[17] affixing a peace symbol to an upside-down American flag,[18] and remaining silent[19] have been found to be sufficiently communicative to constitute symbolic speech.

Conduct, however, has traditionally been subject to governmental control.[20] Consequently, the courts have struggled with differentiating between expressive conduct which constitutes "speech," and should thus be protected by the first amendment, and non-expressive conduct which is "non-speech," and, hence, is not necessarily protected by the first amendment.[21] Because it is almost impossible to draw this line in most symbolic speech cases, the Supreme Court, in *United States v. O'Brien*,[22] effectively eliminated the need to do so by holding that "when 'speech' and 'non-speech' elements are combined in the same course of conduct, a

13. *See* notes 14-19 *infra*. *See also* Note, *First Amendment Protection of Symbolic Speech: Flag Cases Raise the Standard*, 8 LOY. L.A.L. REV. 689 (1975).

14. Spence v. Washington, 418 U.S. 405, 410-11 (1974). The courts have been reluctant to recognize any conduct as speech just because the person engaging in the conduct intends thereby to express an idea. United States v. O'Brien, 391 U.S. 367, 376 (1968). The objective requirement that the message be "likely to be understood" is ascertained by considering both the nature of the symbolic act *and* the circumstances, or context, in which the act was performed. See Spence v. Washington, 418 U.S. 405, 413 (1974). *See generally* Nimmer, *The Meaning of Symbolic Speech Under The First Amendment*, 21 U.C.L.A. L. REV. 29 (1973).

15. Stromberg v. California, 283 U.S. 359 (1931).

16. *See* United States v. O'Brien, 391 U.S. 367 (1968).

17. Tinker v. Des Moines Indep. Comm'ty School Dist., 393 U.S. 503 (1969).

18. Spence v. Washington, 418 U.S. 405 (1974).

19. Russo v. Central School Dist., 459 F.2d 623 (2d Cir. 1972), *cert. denied*, 411 U.S. 932 (1973) (teacher's silence during daily flag salutes).

20. *See, e.g.*, Cantwell v. Connecticut, 310 U.S. 296 (1940) (freedom to believe is absolute, but conduct remains subject to regulation for the protection of society).

21. Conduct which is only incidental to speech, rather than equivalent to speech, is given only limited first amendment protection. Thus, picketing is "speech" only because pickets bear signs on their backs, not because of any special communicative characteristic of picketing, which might alone appear to be loitering. *See* Giboney v. Empire Storage & Ice Co., 336 U.S. 490, 498 (1949). *See also* Cox v. Louisiana, 379 U.S. 536 (1965). But a sit-in constitutes a communicative act *per se*, and may be afforded special first amendment protection. *See* Brown v. Louisiana, 383 U.S. 131 (1966); Edwards v. South Carolina, 372 U.S. 229 (1963). *See generally* Note, *Symbolic Conduct*, 68 COLUM. L. REV. 1091 (1968); *First Amendment Protection, supra* note 13; Comment, *Supression of Symbolic Speech: Anti-Speech versus Non-Speech Interests*, 22 U.C.L.A. L. REV. 969 (1975).

22. 391 U.S. 367 (1968) (burning of a draft card). *See also*, Ely, *Flag Desecration: A Case Study in the Roles of Categorization and Balancing in First Amendment Analysis*, 88 HARV. L. REV. 1482, 1495 (1975).

sufficiently important governmental interest in regulating the non-speech element can justify incidental limitations on [the speech element]."[23] The Court in *O'Brien* articulated the following four part test for determining when a governmental interest is "sufficiently important":

> [1] If it is within the consitutional power of the Government; [2] if it furthers an important or substantial governmental interest; [3] if the governmental interest is unrelated to the suppression of free expression; and [4] if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest.[24]

This test has since become a yardstick to apply against the governmental interest whenever a symbolic speech issue is raised.[25]

Another form of expression recognized by the Court in recent years is involuntary, or compelled, expression. Such forced expression is often termed "forced affirmation of belief" and exists whenever one is "compel[led] . . . to utter what is not in his mind."[26] The Court views this type of compelled expression with great disfavor, and has established clear safeguards against it since the landmark case of *West Virginia Board of Education v. Barnette.*[27] In that case, the Court ruled that a state which compelled school children to participate in a flag salute "invade[d] the sphere of intellect and spirit which it is the purpose of the First Amendment . . . to reserve from all official control."[28] Acknowledging that the flag salute was intended to symbolically convey a message, the Court concluded that the message could not be conveyed through coercion.[29] "If there is any fixed star in our constitutional constellation, it is that no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion, or other matters of opinion or *force citizens to confess by word or act* their faith therein."[30]

---

23. 391 U.S. at 376.

24. *Id.* at 377.

25. *See, e.g.*, Slocum v. Fire & Police Comm'n, 8 Ill. App. 3d 465, 290 N.E.2d 28 (1972) (*O'Brien* analysis used to justify limitation of a symbolic speech interest). *But see* Spence v. Washington, 418 U.S. 405, 414, n.8 (1974) (*O'Brien* analysis not applicable if government's interest suppresses an individual's symbolic expression). The inapplicability of *O'Brien* in some symbolic speech cases is discussed in Ely, *supra* note 22. The district court in *Maynard* also recognized the inapplicability of *O'Brien*, 406 F. Supp. at 1389 n.14. See text accompanying notes 85-86 *infra*.

26. West Virginia State Bd. of Educ. v. Barnette, 319 U.S. 624, 634 (1943).

27. *Id.*

28. 319 U.S. at 642.

29. *Id.* at 632-33.

30. *Id.* at 642 (emphasis added).

Expression which is compelled by the state has only been upheld under very narrow circumstances. The Court in *Barnette* stated that compelled expression could be commanded only upon a showing of "clear and present danger" grounds that are more immediate and urgent than those required to compel silence.[31] The Court concluded that it could not conceive of any circumstances where those grounds might exist.[32] Nevertheless, there have been narrow circumstances under which compelled expression has been permitted. Such circumstances are generally characterized, however, by some significant voluntary act which has placed the individual in a position where he is forced to "speak."[33] Even then, the courts carefully scrutinize the proposed governmental interest requiring the expression to ensure it is "sufficiently important."[34]

B. New Hampshire Law

In 1969, the New Hampshire Legislature decided that its state motto should appear on most non-commercial license plates. The stated purpose for placing the motto on the plates was both to promote appreciation of history, individualism, and state pride and to facilitate the identification of motor vehicles.[35] When the "Live Free or Die" plates appeared in 1971, not everyone accepted them. As early as 1972, in *State v. Hoskin*,[36] challengers of the motto succeeded in reaching the New Hampshire Supreme Court with two questions of law: (1) whether the motto, "Live Free or Die," was included in the "figures or letters" protected by the defacement statute,[37] and (2) whether the display of the motto constituted a "forced affirmation of belief." The challengers lost on both questions. The court reasoned that the display of a motto required by state law carries no implication that the displayers endorse the mes-

---

31. *Id.* at 633-34.

32. *Id.* at 642.

33. For example, the requirement of a loyalty oath, or the wearing of a uniform which conveys a symbolic message, may be viewed as a requirement the individual has imposed upon himself by voluntarily seeking office, or by choosing employment with a particular employer. *See* Slocum v. Fire & Police Comm'n, 8 Ill. App. 3d 465, 290 N.E.2d 28 (1972). Loyalty oaths, however, are usually sustained on grounds that they are constitutionally mandated, or are necessary if the government is to survive. *See, e.g.*, American Communications Ass'n v. Douds, 339 U.S. 382 (1950); Biklen v. Board of Educ., 333 F. Supp. 902 (N.D.N.Y. 1971), *aff'd* 406 U.S. 951 (1972).

34. *E.g.*, Biklen v. Board of Educ., 333 F. Supp. 902 (N.D.N.Y. 1971).

35. 97 S. Ct. at 1436. *But see* Appendix, *supra* note 2, at 44 (a legislator had explained the purpose of the motto was to advertise that New Hampshire residents believe in fighting for democracy).

36. 112 N.H. 332, 295 A.2d 454 (1972).

37. See note 2 *supra*.

sage.[38] Hence, the court in *Hoskin* construed forced affirmation of belief as requiring the state to "place the citizen in the position of either appparing to, or actually, 'asserting as true' the message."[39]

III. The Supreme Court Opinion

The United States Supreme Court found "more appropriate First Amendment grounds" to affirm the decision than the symbolic speech grounds relied upon by the lower court. The lower court had been easily convinced that the Maynards' conduct contained the requisite elements for symbolic speech. The district court inferred the requisite intent to communicate from the use of reflective red tape to mask the motto, which not only called attention to the fact that the motto was obscured, but also communicated Maynards' disagreement with it.[40] That the Maynards' message was likely to be understood was evident from the general knowledge among New Hampshire citizens of the motto on their license plates and the opposition to the motto by individuals such as the Maynards.[41]

The Supreme Court, however, was skeptical as to the existence of symbolic speech because Maynard had requested plates without the motto.[42] This request evidences that he wanted to avoid prosecution[43] rather than communicate any message.[44]

---

38. The court in *Hoskin* analogized the motto on license plates to United States money which bears the motto "In God We Trust." Carrying such currency does not demonstrate a belief in the motto. It also pointed out that nothing precluded disagreement with the motto provided the methods used did not obscure the number plates. 294 A.2d at 457.

39. 97 S. Ct. at 1438-39 (Rehnquist, J., dissenting).

40. 406 F. Supp. at 1387. Also of interest is the following testimony taken at a hearing before the district court:

> Q. [Direct Examination] Can you tell the Court why you used this bright orange or red flourescent type tape or reflective tape?
>
> A. [Mr. Maynard] The reason for it is that people will recognize what I am doing, which is very effective. A lot of people stop me. And one person says "You can't do that. That's against the law." I says "Fortunately, I was given permission by the Federal Court in a temporary injunction against the State." And here I was able to converse with him and express my beliefs and my reason for doing so.
>
> Q. Has this taping caused any outbursts or brawls or anything of that nature?
>
> A. No, they haven't. I can see in any rear-view mirror, people stop at a stop sign and are pointing at it and bringing everybody else's attention to it, and that tells you that people are taking notice of the tape.

Appendix, *supra* note 2, at 29.

41. *See* 406 F. Supp. at 1387 n.11.

42. The Maynards had requested commercial plates, which do not bear the motto, but their request was denied. Appendix, *supra* note 2, at 36. They also requested in their complaint that the Commissioner of Department of Motor Vehicles be required to issue them plates without the state motto in the event New Hampshire succeeded on the merits. *Id.* at 10.

43. The district court suspected that "Mr. Maynard had not conceived of the symbolic

The more appropriate grounds the Court found were the Maynards' first amendment right of freedom of thought. Included within this broad freedom was the "right to speak and the right to refrain from speaking," such rights being "complementary components of the broader concept of 'individual freedom of mind.'"[45] The Court relied on *Miami Herald Publishing Co. v. Tornillo*[46] to illustrate the fact that all first amendment rights are really composed of positive rights and negative rights, one the complement of the other.[47] In *Tornillo,* the Court held that freedom of the press included the fundamental right to decide both what to print and what not to print.[48]

The controlling authority recognizing a first amendment right of freedom of thought, and hence the right to speak and the right to refrain from speaking, was *West Virginia State Board of Education v. Barnette.*[49] The Court found that Maynard, like the school children in *Barnette,* was compelled by the state to be an instrument to foster an ideological message which he found abhorrent. That is, Maynard was forced to affirm symbolically the motto "Live Free or Die," which he did not believe.[50]

In dissent, Justice Rehnquist questioned whether the mere carrying of a license plate embossed with the words "Live Free or Die" was an affirmation of any kind. He argued that before compelled expression can be found, one must first find that expression has taken place. The key issue, according to Justice Rehnquist was "whether [the Maynards] in displaying . . . state license tags . . . would be considered to be advocating political or ideological views." "The State," he argued, "has not forced [the Maynards] to 'say' anything."[51] Justice Rehnquist concluded his argument by stating

---

speech approach until very competent expert counsel called his attention to it." Appendix, *supra* note 2, at 62-63.

44. *See* 97 S. Ct. at 1434 n.10.

45. 97 S. Ct. at 1435.

46. 418 U.S. 241 (1974).

47. For example, the first amendment includes a right to publish and not to publish, a right to practice religion and not to practice religion, a right to assemble and not to assemble, a right to speak and not to speak. All of these rights are included in the broader first amendment concept of freedom of thought. 97 S. Ct. at 1435.

48. 418 U.S. 241 (1974).

49. 319 U.S. 624 (1943). See notes 27-30 *supra* and accompanying text.

50. The Court acknowledged that compelling the "*affirmative* act of a flag salute involved a more serious infringement upon personal liberties than the *passive* act of carrying the state motto on a license plate." But this difference was said to be only "one of degree." Even though the act was passive, it nonetheless made Maynard's automobiles appear "as a *mobile billboard* for the State's ideological message." 97 S. Ct. at 1435. (emphasis added).

51. *Id.* at 1438. Compare this issue as stated in the dissent to the issue as framed by the majority: "We are . . . faced with the question of whether the State may constitutionally

the Court's holding would lead to startling and unacceptable results, such as a finding that the motto "In God We Trust," which appears on all United States currency, would violate an atheist's first amendment rights.[52]

The majority concluded their analysis by balancing the identified first amendment interest—compelled expression—against the state's countervailing interests. Citing *United States v. O'Brien*[53] as authority for using a balancing test,[54] the Court found that the state's interests were insufficiently weighty.[55] Thus, in affirming the district court,[56] the Court concluded "that the State of New Hampshire may not require [the Maynards] to display the state motto upon their vehicle license plates."[57]

---

require an individual to participate in the dissemination of an ideological message by displaying it on his private property in a manner and for the express purpose that it be observed and read by the public." *Id.* at 1434-35.

52. *Id.* at 1439. Compare note 38 *supra*. Justice Rehnquist also criticized the Court for failing to explain, after pointing out the "obvious differences" between this case and *Barnette*, why the same result should be obtained. The "conclusory words" used by the Court, such as *foster* and *instrument*, were felt by Justice Rehnquist to be "barren of analysis." 97 S. Ct. at 1438. To illustrate, Justice Rehnquist hypothesized that "were New Hampshire to erect a multitude of billboards, each proclaiming 'Live Free or Die,' and tax all citizens for the cost of erection and maintenance, clearly the message would be 'fostered' by the individual citizen-taxpayers and just as clearly those individuals would be 'instruments' in that communication." *Id.*

Justice Rehnquist also assailed the Court for failing to discuss State v. Hoskin, 112 N. H. 332, 295 A.2d 454 (1972), wherein the New Hampshire Supreme Court expressly held that displaying a motto on a license plate was not a forced affirmation of belief, nor in violation of the holding in *Barnette*. *Id.* at 1439. See notes 36-39 *supra* and accompanying text.

53. 391 U.S. 367 (1968).

54. 97 S. Ct. at 1436.

55. The state argued that the motto facilitates identification of passenger vehicles. The record revealed, however, that different configurations of numbers and letters made the plates "readily distinguishable from other types of plates, even without reference to the state motto." *Id.* Hence, even though the state had a legitimate purpose—to identify vehicles—that purpose could not be achieved through regulations that "broadly stifle fundamental liberties" when more narrow means could be used. *Id., quoting* Shelton v. Tucker, 364 U.S. 479, 488 (1960).

The state's other objective—to promote the message of "appreciation of history, individualism and state pride"—was found to be "not ideologically neutral." While such interests could be pursued by the state in various ways, they could not be pursued in a way that violated "an individual's first amendment right to avoid becoming the courier for such ideological message." *Id.*

The Court in *Maynard* declined to comment on what methods the state could use to legitimately pursue such interests. At least one method New Hampshire has subsequently pursued is to imprint the state motto on all state stationery. Letter from Robert V. Johnson, II, Assistant Attorney General of New Hampshire, to Bryant R. Gold (Sept. 26, 1977).

56. In granting the injunctive relief, the district court had found that New Hampshire's interests did not meet the third and fourth standards of the *O'Brien* test for "sufficiently important" governmental interests. 406 F. Supp. at 1388. See note 24 *supra* and accompanying text.

57. 97 S. Ct. at 1436. A very significant, but collateral, issue confronting both the

IV. ANALYSIS

A. *Compelled Expression*

The principal dividing line between the majority and the dissent in *Maynard* was whether the case involved any form of compelled expression. The majority assumed there was without really justifying why. The dissent argued there was not, relying on the rationale of *State v. Hoskin.*[58] This conflict rests on whether one must apply an objective or subjective test to determine whether compelled expression exists in a given situation.

The dissent's approach to find compelled expression was manifestly *objective.* Justice Rehnquist argued the citizen must *appear* to be asserting the truth of the message. The issue, according to the dissent, was whether the citizen "would be considered to be advocating political or ideological views."[59] This language leaves little doubt that the dissent's approach is to look at the citizen's conduct from without, as a reasonable observer, and to make an objective judgment as to whether it appears to others that the citizen is asserting a particular message.

The majority's approach, on the other hand, was impliedly *subjective.* The "essence of [the] objection" was that the Maynards considered the motto to be "morally, ethically, religiously and politically abhorrent."[60] Hence, the majority was not concerned with what others thought relative to the expressive content of the Maynards' compelled act. If the Maynards believed they were forced to make an expression, that sufficed.

While an objective approach is perhaps useful in finding "symbolic speech," at least insofar as that term has come to mean

---

district court and the Supreme Court in *Maynard* was the applicability of the doctrine of equitable restraint as ennunciated in Younger v. Harris, 401 U.S. 37 (1971). *Younger* required that a federal court should not enjoin a pending state criminal prosecution unless there were unusual circumstances, such as bad faith or harassment on the part of a state. In addition, if the plaintiff could not show irreparable injury, then declaratory relief should also be barred. The Court found an injunction justified because of the successive prosecutions to which the Maynards had been exposed and the threat of future prosecutions which would, in effect, deprive them of the use of their automobile, a necessity to perform the ordinary tasks of daily life. 406 F. Supp. at 1384-85, 97 S. Ct. at 1433-34.

Justice White, joined by Justice Blackmun and Justice Rehnquist dissented from the granting of an injunction. They argued that a declaratory judgment must first be issued. 97 S. Ct. at 1436. For an excellent summary of the *Younger* doctrine, and the line of cases which have shaped that doctrine, see C. WRIGHT, LAW OF FEDERAL COURTS, § 52A (3d ed. 1976). For a discussion of how the *Younger* doctrine applies to the *Maynard* case, see Comment, *Compelled Expression: Maynard v. Wooley,* 28 ME. L. REV. 531, 533 n.18 (1976).

58. 112 N.H. 332, 295 A.2d 454 (1972).

59. 97 S. Ct. at 1438-39 (Rehnquist, J., dissenting).

60. *Id.* at 1434.

a voluntary, affirmative, expressive act,[61] it becomes suspect when applied to finding "compelled expression." One of the goals of the Bill of Rights, including the first amendment, is to protect individual expression from governmental control.[62] Applying an objective test to find compelled expression defeats that goal, for it forces an individual to express a view not his own. This situation clearly invades "the sphere of intellect and spirit which it is the purpose of the First Amendment . . . to reserve from all official control."[63]

Care must be exercised, however, to distinguish compelled symbolic expression from compelled acts which have no expressive content.[64] To make this distinction, an objective test is correctly applied first to ascertain if the state, or governmental agency, intended to communicate symbolically a message. If it did, then a subjective test is properly applied to determine if the individual citizen was compelled to communicate that message. If it did not, then there exists no expression to compel. In other words, "expression" is a prerequisite of "compelled expression,"[65] and "expression," if symbolic in nature, is properly found by using an objective test.[66] Expression that is compelled, however, must be

61. *See* notes 14-19 *supra* and accompanying text.

62. "The very purpose of a Bill of Rights was to withdraw certain subjects from the vicissitudes of political controversy, to place them beyond the reach of majorities and officials and to establish them as legal principles to be applied by the courts." West Virginia State Bd. of Educ. v. Barnette, 319 U.S. 624, 638 (1943).

63. *Id.* at 642.

64. For example, the state can compel one to place a license plate on his car, or a number on his house, and can even compel the manner in which his house is constructed. But such acts are not intended by the state to communicate any message. They are intended only to protect the general welfare of the citizens.

65. Some of the language used by Justice Rehnquist in dissent, indicates the nature of his objection was that an objective test had not been used first. "The State . . . has not forced the Maynards to communicate ideas with nonverbal actions reasonably likened to 'speech,' such as wearing a lapel button, promoting a political candidate or waiving a flag as a symbolic gesture. . . . What the Court does not demonstrate is that there is any 'speech' or 'speaking' in the context of this case." 97 S. Ct. at 1438. It is equally clear, however, from Justice Rehnquist's language that he would also apply an objective test to determine if Maynard "would be considered to be advocating political or ideological views." *Id.*

66. *See* note 14 *supra.* Not only is "expression" found using an objective test, which requires a communicator who intends to express an idea that is *likely* to be understood, but the Court in *Maynard* requires the "expression" to be "ideological" in nature. *Id.* at 1435. This suggests that what constitutes an "ideological message" is likewise an objective determination.

This ideological requirement sounds like that proposed by the political philosopher, Alexander Meiklejohn. Meiklejohn argues that one must distinguish between two kinds of expression: (1) political expression, which is protected by the first amendment, and (2) private expression, non-political in character, which does not fall within the ambit of the first amendment, but falls instead under the fifth amendment, and may be abridged. *See* A. MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT 38 (1948).

found subjectively if the guarantees of the first amendment are to have any real significance.

In *Maynard,* there can be little doubt that the state intended to communicate a message. Indeed, the express purpose of placing the motto on the plates was to promote appreciation of history, individualism and state pride,[67] *i.e.,* to communicate or express that message to all who would see and read the motto on the plates.[68] If that message were not likely to be understood, then the legislature would not have gone to the trouble and expense of embossing it on the license plates. Hence, the state's conduct plainly meets the requirements of symbolic speech.[69] Therefore, the majority properly applied a subjective test to find that such expression had been compelled in violation of the Maynards' first amendment rights.

B. Compelled Expression v. Symbolic Speech

The objective-subjective approach suggested by the *Maynard* decision represents a useful technique for the Court to use hereafter whenever a first amendment freedom of expression interest is implicated. This technique, or test, need not be limited to solely "compelled expression," but may be applied to either compelled expression or symbolic speech by combining the two tests heretofore used.

The combined test, simply stated, is:

(1) Objectively determine if a "symbolic message" was intended to be communicated. This requires (a) an intent to communicate symbolically on the part of the communicator,[70] and (b) the likelihood

---

67. *See* note 35, *supra* and accompanying text.

68. One of the arguments advanced by the appellants in refuting Maynard's claim of symbolic speech was that the symbolism of taping over the motto "Live Free or Die" has different meanings to different people. Thus, the argument goes, it cannot be said that any one message was intended, when so many messages are possible. Brief for Appellants at 26, 97 S. Ct. 1438 (1977). *See also* 496 F. Supp. at 1386 n.10. This same argument could be advanced with respect to the symbolism the state hopes to advance by forcing display of the motto. That is, it is impossible to tell what the state's message is because each individual attaches his own meaning to the motto. However, such an argument only makes the state's case weaker, for if the state intended a broad choose-your-own-meaning message, then a form of expression exists. And if, after choosing-his-own-message, the individual subjectively believes he is forced in that expression, then he has good grounds for finding compelled expression.

69. See text accompanying note 14 *supra.*

70. To "communicate symbolically" means to use an individual, or his private property, in such a fashion that he appears as a symbol that conveys a desired message. Such appearance will generally require a high level of individual participation in the form of action, as opposed to a very low, or minimal, level of participation. Thus, a taxpayer required to pay taxes for the construction of a stationary billboard which expresses the state's ideological viewpoint, see note 52 *supra,* would probably not be construed as the "symbol" of the state's

that the message will be understood by the communicatee. If so,

(2) Determine the identity of the communicator.

(a) If the communicator is an individual, then a positive freedom of expression interest has been found—heretofore called "symbolic speech."

(b) If, on the other hand, the communicator is the government, then subjectively determine if the complainant is "forced" or "compelled" to express the communicator's message. If so, then a negative freedom of expression interest has been found—heretofore termed "compelled expression."

This approach may be represented diagramatically as shown in Figure 1.

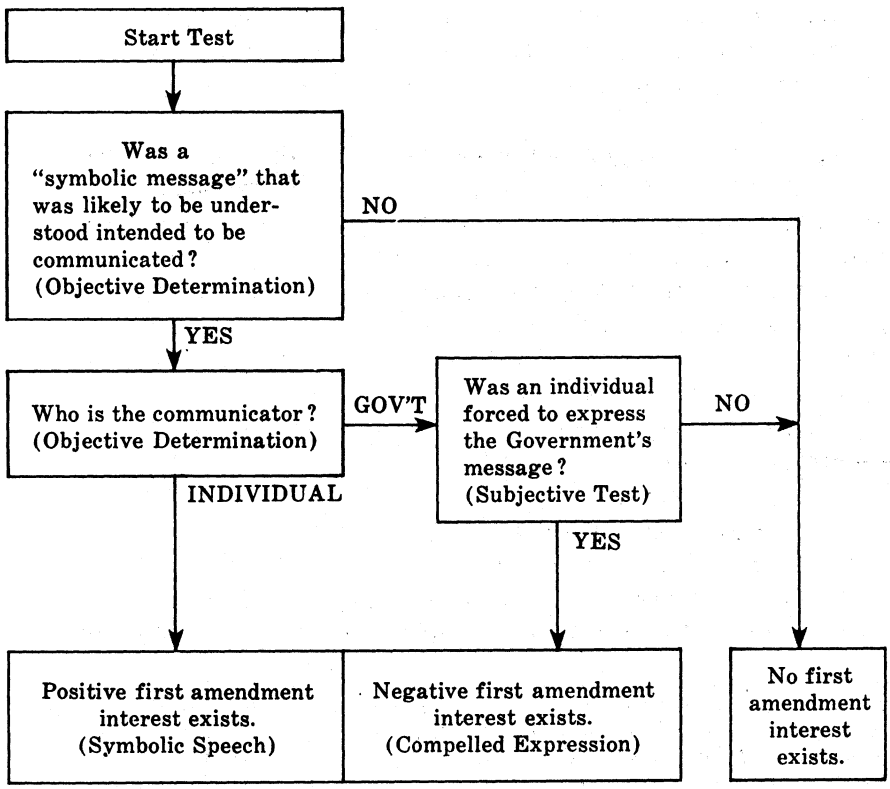


Figure 1. Combined Test for Identifying a Positive or Negative First Amendment "Expression" Interest.

message because (1) his individual participation in such activity is minimal, involving only the payment of money, which few would observe, *see Compelled Expression, supra* note 57, at 542-43, and (2) the "payment of money" in this context would probably not be considered the "use of private property," since it was converted to public funds by way of the tax. *See, e.g.*, Joyner v. Whiting, 477 F.2d 456, 461 (4th Cir. 1973) (state agency may spend public

The combined test, while representing little that is new, provides at least two significant advantages over the isolated tests presently employed to find symbolic speech or compelled expression. First, the combined test eliminates the need to initially identify a specific positive or negative interest. Heretofore, such identification could be extremely important due to the inherently inconsistent nature of symbolic speech and compelled expression. For example, the key element in the test presently used to find symbolic speech is the showing of an *intent* to communicate; there must be both a communicator and a receiver, the one intending to communicate to the other.[71] On the other hand, the test for compelled expression requires a finding that the communication or expression is forced, or done *without intent.*[72] Thus, to argue both symbolic speech and compelled expression *for the same act* is to argue inconsistent theories. The stronger the case for one theory, the weaker the case for the other.

The facts of the *Maynard* case illustrate this potential inconsistency. Arguably, Maynard's claims of "symbolic speech" and "compelled expression" were based on two separate acts. Maynard's claim of compelled expression was directed at the *state's act* of requiring him to carry or display the motto to which he disagreed. His claim of symbolic speech was directed to *his act* of covering over the motto, thus voicing symbolically his opposition to it. Such positions, of course, are not inconsistent.[73]

On the other hand, one could argue that the state's act of requiring Maynard to carry the motto was so closely related to Maynard's act of covering over the motto that there was effectively only one act involved. If so, an inconsistency would exist for it would imply Maynard *intended to* carry the motto to express opposition to it, but at the same time he *did not intend to* carry the motto because to do so was a form of compelled expression. It was precisely

---

money to *advocate* a controversial position even though the execution of the position advocated would be unconstitutional). *But cf.* Stern v. Kramarsky, 375 N.Y.S.2d 235 (1975) (state agency may not use public funds to advocate passage of Equal Rights Amendment); Bonner-Lyons v. School Comm. of City of Boston, 480 F.2d. 442 (1st Cir. 1973) (school committee may not advocate busing policies by sending its views home with school children in the form of notices and letters unless those with differing viewpoints are afforded the same opportunity).

71. See note 14 *supra* and accompanying text.

72. See note 26 *supra* and accompanying text.

73. It is as though Maynard had said: I don't want to be the carrier for the state's ideological message. To force me to carry their message is an act of compelled expression, and I have a first amendment right protecting me from such compelled expression. But, inasmuch as I have to carry the message anyway, I intend to symbolically express my opposition to it through my separate act of covering it over wiith bright reflective tape.

this type of inconsistency which persuaded the majority in *Maynard* to infer that Maynard's act of covering over the motto might not be protected symbolic speech.[74]

Thus, counsel must initially decide whether to argue only that interest which comes closest to fitting the facts, with the risk of choosing the wrong interest and losing, or whether to argue inconsistent theories, with the risk of appearing confused before the fact finder.

The combined test eliminates these problems by ending the necessity of choosing which party was the communicator prior to finding intent. Intent must still be found, but the party in whom it is found is not critical. Thus, counsel need not concern himself with which theory—positive or negative—to initially plead.[75] All he need do is plead the underlying theory which encompasses both—first amendment freedom of expression.

Second, as evident from the preceeding paragraph, the combined test brings out and emphasizes the common interest that the first amendment, including the speech clause, protects—freedom of expression. As the Court in *Maynard* aptly pointed out, "[a] system which secures the right to proselytize religious, political, and ideological causes must also guarantee the concomitant right to decline to foster such concepts."[76] Thus, whether an individual desires to "say" something,[77] and is encumbered or prohibited in that expression through governmental regulation, or whether he desires not to "say" anything, but is compelled by the government to express a view not his own, the protections of the first amendment ought to provide equal immunity from such governmental actions.[78]

---

74. *See* 97 S. Ct. at 1434 n.10.

75. "Silence" is perhaps the best example of an act which would be either "symbolic speech" or "compelled expression." For example, a school teacher may wish to remain silent during the daily flag ceremony, thereby communicating to the students a message of dissatisfaction with what the obligatory flag salute symbolizes. Alternatively, the teacher may not wish to communicate any message at all, but may simply not want to be compelled in an act of involuntary expression. Under these facts, a federal court has held that the teacher's first amendment rights were violated, but whether the holding was based on positive grounds or negative grounds, or both, is unclear. *See* Russo v. Central School Dist., 469 F.2d 623 (2d Cir. 1972).

76. 97 S. Ct. at 1435.

77. "Say," as used here, refers to both verbal expression and symbolic expression. For an argument that both types of expression should be treated equally under the first amendment, see Nimmer, *supra* note 14.

78. *See generally* EMERSON, *supra* note 10; Grey, *Do We Have an Unwritten Constitution?* 27 STAN. L. REV. 703 (1975); Kurland, *The Privileges or Immunities Clause: "Its Hour Come Round at Last?"* 1972 WASH. U.L.Q. 405.

C. *The Government's Interest*

Once a first amendment interest has been identified, it then becomes necessary to determine if that interest will prevail over any countervailing government interest. In *O'Brien,* the Court prescribed a coherent balancing technique that could be used to make that decision.[79] Because *O'Brien* involved a positive speech right—symbolic speech—the balancing test it described has generally been thought to be applicable only to symbolic speech.[80] That such a narrow application of the balancing test was not intended by the Court was made clear when the Court in *Maynard* cited *O'Brien* as authority for balancing a negative speech right, compelled expression, against the government interests.[81]

Despite this apparent applicability of the *O'Brien* balancing test to both positive and negative speech rights, the Court has recognized that a balancing technique is not appropriate in certain situations.[82] In these situations, instead of balancing, the Court considers only whether the expression is of a type which will incite, or is likely to incite, imminent lawless action.[83] If the expression tends to incite imminent lawless action, it is said to fit into a category of unprotected expression; if it does not, it is absolutely protected. This approach has been termed "categorization."[84]

---

79. See text accompanying note 24 *supra.* Not everyone agrees that "balancing" is a legitimate method to be used in the area of first amendment freedoms. Balancing, it is argued, is too inconsistent and unpredictable due to the prejudices and biases of those performing the balance. Balancing thus becomes "nothing more than a way of rationalizing preformed conclusions." EMERSON, *supra* note 10, at 718. *See also* Konigsberg v. State Bar of California, 366 U.S. 36, 61 (1961) (Black, J., dissenting) ("the men who drafted our Bill of Rights did all the 'balancing' that was to be done"). *Contra,* Grey, *supra* note 78.

80. *See, e.g., First Amendment Protection, supra* note 13, at 700; *Compelled Expression, supra* note 57, at 532.

81. 97 S. Ct. at 1436. *See also* Ely, *supra* note 22 at 1484 n.11 (the scope of the *O'Brien* balancing test was never intended to be restricted to symbolic speech).

82. *See, e.g.,* Cohen v. California, 403 U.S. 15 (1971); Brandenburg v. Ohio, 395 U.S. 444 (1969) (per curiam); Ely, *supra* note 22, at 1490-93.

83. Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (per curiam).

84. *See* Ely, *supra* note 22 at 1484. The categorization approach absolutely protects all first amendment expression except that which will produce, or is likely to incite, imminent lawless action. The term "categorization" may be somewhat of a misnomer, for it suggests absolute protection for interests which do not fall within a "narrowly defined category" of unprotected interests and infers that no balancing ever takes place. However, the court must still attach practical meaning to the terms "produce," "likely to incite," "imminent" and "lawless," and such a process may invoke the same policy considerations as are involved in "balancing."

For an interesting historical perspective concerning "categorization" and the conflicting views as to what typs of expression fall into the protected category, see Gunther, *Learned Hand and the Origins of Modern First Amendment Doctrine: Some Fragments of History,* 27 STAN. L. REV. 719 (1975).

At first glance, it may appear that "balancing" and "categorization" are mutually exclusive approaches for testing the validity of a first amendment interest. Professor John Hart Ely argues they are not. Rather, he argues that the categorization approach is merely an extension of the *O'Brien* balancing test, the former to be used only when one of the latter's criteria fails.[85]

This approach, of "switching over" from *O'Brien* balancing to categorization, represents a practical and valuable technique for testing the validity of a first amendment speech interest, whether that interest be positive or negative. The technique maintains the advantages of a coherent balancing approach, while still providing absolute protection in certain instances.

The essence of Professor Ely's "switching-over" approach, modified to include both positive and negative speech rights, may be summarized as follows:[86]

(1) Apply the reduced *O'Brien* criteria[87] to the identified first amendment interst. That is,

[1] Does the governmental control of expression further an important or substantial governmental interest? If yes,

[2] Is the governmental interest unrelated to the control of individual expression? If yes,

[3] Is the incidental restriction on the first amendment freedom no greater than is necessary to the furtherance of that interest? If yes, then the governmental regulation must give way to the first amendment interest. Otherwise, the governmental regulation is justified.

(2) If the answer to criterion [2] is negative, *i.e.*, if the governmental regulation is related to the control of expression, then "switch over" to the categorization approach, and determine whether the expression is the type which will produce, or is likely to incite, imminent lawless action. If yes, the governmental control of expression is justified. Otherwise, the first amendment interest is absolutely protected.

Diagramatically, this modified "switching-over" approach may be represented as illustrated in Figure 2.

---

85. *See* Ely, *supra* note 22, at 1484, 1497-98.

86. *See generally* Ely, *supra* note 22.

87. Professor Ely eliminated *O'Brien's* first criterion as superfluous. See Ely, *supra* note 22, at 1483 n.10.

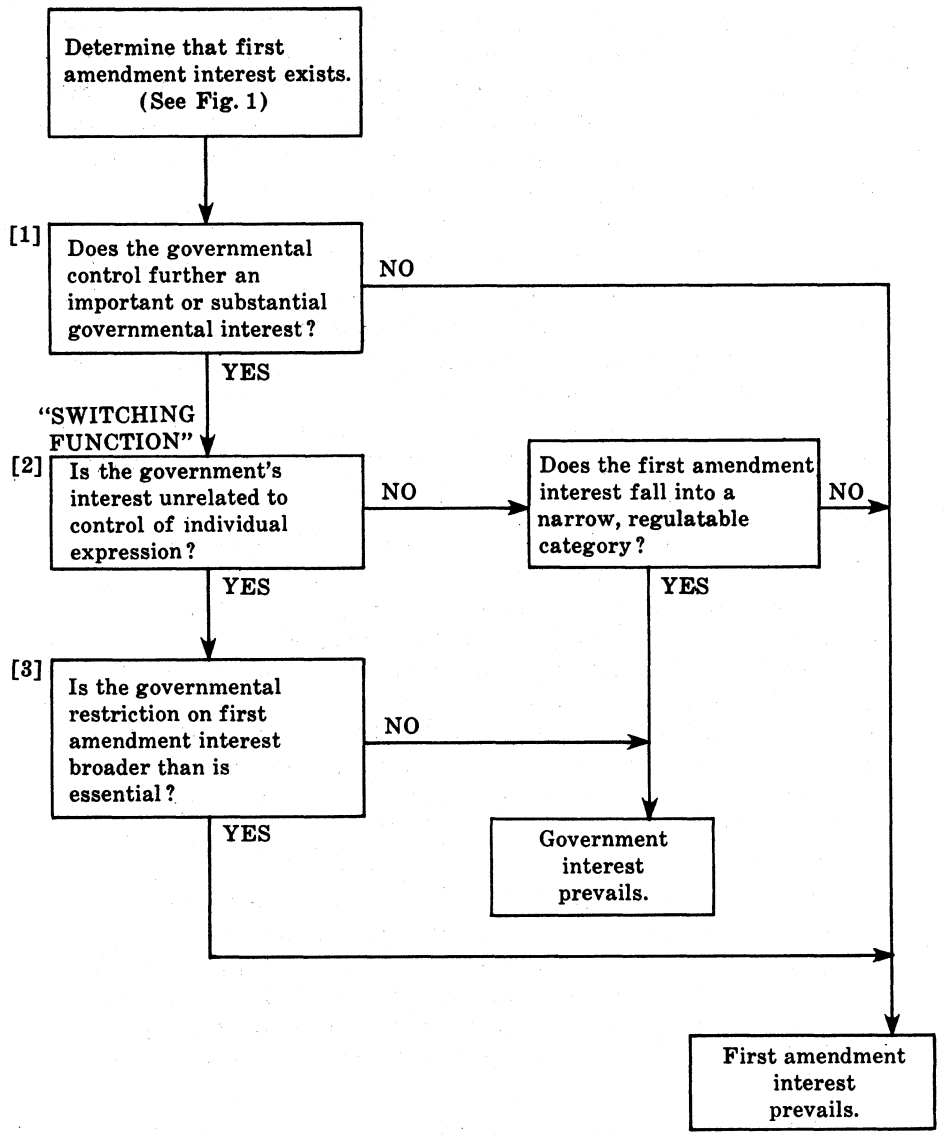


Figure 2. Diagramatic Representation of a Proposed Framework for the Resolution of Questions, concerning Both Positive and Negative Free Expression Interests.

An analysis of Professor Ely's proposed method reveals that the "switching function" is little more than a question as to whether the end is legitimate. An illegitimate end is defined as one which is related to the control of expression, whether it be the suppressing of expression or the forcing of expression. Thus, if Maynard's act of covering over the motto on the license plate is viewed as a positive

right—symbolic expression—then the suppressing of that expression by making it inferior to the government's expression is an illegitimate end.[88] On the other hand, if forced display of the motto is viewed as a negative right—compelled expression—then by definition it is clearly control of expression and is also an illegitimate end. Thus, by either approach, the analysis is "switched over" to the categorization approach.[89]

If the interst is *not* "switched over," then the last of the *O'Brien* criteria is reached. This criterion relates to whether the restriction is no greater than essential and requires a balancing to determine the necessity of the restrictive means selected by the government. This process, argues Professor Ely, should be a "serious balancing," not just a question of whether there is an alternative means capable of serving the government's interest as efficiently as it is served by the regulation under analysis. But in many instances, such as in the *Maynard* case, the balancing question need never be reached because the issue is "switched over" to the categorization approach.

In *Maynard,* the act of covering over the motto—symbolic expression—was "lawless" only because it violated New Hampshire's license plate defacement statute.[90] However, such a concept of "lawless" is hardly consistent with the policy underlying the categorization approach.[91] In addition, an argument that Maynards' conduct would produce, or was likely to incite, imminent lawless action can quickly be refuted by the record. No "imminent lawless action" of the type contemplated under the categorization approach resulted,[92] and Maynard had engaged in his "lawless" conduct for almost a full year before his first citation.[93]

Similarly, the state's act of requiring Maynard to display the motto against his will—compelled expression—does not fit within any of the "narrowly defined categories" where forced expression would be justified. Generally, forced expression has been constitutionally upheld only where there exists a constitutional mandate

88. *See id.* at 1506-08.

89. This analysis demonstrates that any attempt to compel expression will always be "switched over" to the categorization approach. Regulation of "symbolic speech," on the other hand, will not always be "switched over" because often the regulation is aimed only at preventing interruption of a legitimate activity rather than suppression of the symbolic speech. *Id.* at 1498-1500.

90. See note 3 *supra.*

91. Categorization protects all activities except those that would result in direct incitement to, or direct advocacy of, actions which are dangerous and disruptive to society. *See generally* Gunther, *supra* note 84.

92. See note 40 *supra.*

93. *See* 97 S. Ct. at 1431-32.

that it be done, as where a public official or servant is required to affirm his loyalty to a certain cause.[94] There is no such constitutional mandate in the *Maynard* case. Thus, the Maynards' right to be free from compelled expression must be upheld.

V. Conclusion

The Supreme Court in *Wooley v. Maynard* correctly held that New Hampshire could *not* force the Maynards to carry publicly an ideological message, to which they were opposed, on their private property. The Court relied on the constitutionally protected right "not to speak" in arriving at this result. In contrast, the lower court held for the Maynards on "symbolic speech" grounds. Both of these rights come within the ambit of the first amendment's freedom of expression, each being the complement of the other. Whether *Maynard* is viewed as involving "compelled expression" or "symbolic speech," it clearly concerned freedom of expression—the freedom to express one's own ideas and/or the freedom to not express another's ideas. Under either analysis, the same result must attach if that first amendment freedom is to have any real significance.

The Court's analysis of *Maynard,* while directed only at compelled expression, suggests a method of treating both symbolic speech and compelled expression so that both are analyzed in the same way under the first amendment. The existence of these free speech interests can be identified using a single freedom of expression test without the need of any preliminary classification. In addition, once the freedom of expression interest has been identified, a single test may be used to measure the government's countervailing interest to determine which interest will ultimately prevail.

Bryant R. Gold

94. See note 33 *supra* and accompanying text.

Zacchini v. Scripps-Howard Broadcasting Company: Media Appropriation, the First Amendment and State Regulation

In August and September of 1972, Hugo Zacchini performed his "human cannonball" act at the Geauga County Fair in Burton, Ohio. The act consisted of being shot from a cannon into a net 200 feet away. Zacchini performed in an open grandstand for the pleasure of anyone attending the fair; no separate admission was charged. A free-lance news reporter for Scripps-Howard Broadcasting Company attended the fair with a small movie camera. Zacchini asked the reporter not to film his act. The reporter complied, but returned the following day and, under orders from the producer of the company's news program, filmed Zacchini's performance. The film clip, lasting about fifteen seconds,[1] along with favorable commentary,[2] was broadcast on Scripps-Howard's eleven o'clock news program. Zacchini brought an action for invasion of privacy against Scripps-Howard claiming that the defendant appropriated Zacchini's professional talents for its own use without his consent. The trial court granted defendant's motion for summary judgment.[3] The Ohio Court of Appeals reversed,[4] holding that the complaint stated a cause of action for conversion and for infringement of a common law copyright, and stating that the defendant had no first amendment privilege to show plaintiff's entire act on its news program without compensating Zacchini for financial injury. The Ohio Supreme Court reversed the court of appeals,[5] recognizing petitioner's right to compensation for publicity value, but granting immunity to Scripps-Howard on the basis of freedom of the press as guaranteed by the first amendment. In *Zacchini v. Scripps-Howard Broadcasting Co.*,[6] the United States Supreme Court held that the first and fourteenth amendments of the United States Constitution do not

1. The record is not explicit on exactly what was shown. Zacchini v. Scripps-Howard Broadcasting Co., 97 S. Ct. 2849, 2859 n.1 (1977) (Powell, J., dissenting). For the significance of the content of the broadcast, see notes 53-59 *infra* and accompanying text.

2. The script of the commentary accompanying the film clip read as follows:

> This . . . now . . . is the story of a *true spectator* sport . . . the sport of human cannonballing . . . in fact, the great Zacchini is about the only human cannonball around these days . . . just happens that where he is is the great Geauga County Fair, in Burton . . . and believe me, although it's not a *long* act, it's a thriller . . . and you really need to see it *in person* . . . to appreciate it.

97 S. Ct. at 2851 n.1 (emphasis in original).

3. *See* Zacchini v. Scripps-Howard Broadcasting Co., 47 Ohio St.2d 244, 351 N.E.2d 454, 456 (1976).

4. *See id.* at 456-57.

5. *Id.* at 462.

6. 97 S. Ct. 2849 (1977).

immunize the news media from damage suits when a broadcaster, as part of a news program, broadcasts a performer's entire act without his consent.

I

The right of publicity is based on an individual's right to be free from the appropriation of his name or likeness by another for the other's financial benefit.[7] Courts initially considered a person's name or likeness to be his property and premised protection of freedom from appropriation on traditional property theories.[8] However, the right to be free from appropriation has also been considered to be a part of the general right of privacy.[9] As the courts began to focus on the privacy right, the protection afforded freedom from appropriation came to be based on protection of plaintiff's *feelings* rather than plaintiff's *property*.[10] In order to recover under the privacy

---

7. The interest protected by appropriation has been said to be "the exclusive use of [the actor's] own identity, insofar as it is represented by [the actor's] name or likeness, and insofar as the use may be of benefit to [the actor] or to others." RESTATEMENT (SECOND) OF TORTS § 652C, Comment a (1977).

8. Pavesich v. New England Life Ins. Co., 122 Ga. 190, 50 S.E. 68 (1905) (plaintiff had a property right in his picture which was appropriated by the defendant in defendant's advertising campaign); Munden v. Harris, 153 Mo. App. 652, 134 S.W. 1076 (1911) (defendant used plaintiff's picture to advertise jewelry; defendant liable for damages because plaintiff had a property right in the value of his photograph); Edison v. Edison Polyform Mfg. Co., 73 N.J. Eq. 136, 67 A. 392 (1907) (Thomas A. Edison, a public figure, granted an injunction against appropriation of his property right in his photograph which defendant used to advertise medicine, even though the formula for the medicine had been purchased from Edison). *See generally* Gordon, *Right of Property in Name, Likeness, Personality and History*, 55 NW. U.L. REV. 553 (1960).

9. The right of privacy became a legal right through the influence of a law review article by Samuel D. Warren and Louis D. Brandeis. Warren and Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890). Warren and Brandeis were concerned with intrusions by the press into the private lives of citizens. *Id.* at 196. The authors, however, recognized the distinction between a person who seeks to protect his private life from public notoriety and a person who seeks notoriety for some aspects of his life and, by so doing, waives his right to privacy. *Id.* at 215. Performers and political candidates fall into the latter category. Within the eighty years since the article was written, the right of privacy has been recognized in at least five states by statute, and in at least twenty-six others by judicial decision. *See* Prosser, *Privacy*, 48 CALIF. L. REV. 383, 386-88 (1960).

Dean Prosser divided the right of privacy into four separate torts, each of which protects separate interests:

1. Intrusion upon the plaintiff's seclusion or solitude, or into his private affairs.
2. Public disclosure of embarrassing private facts about the plaintiff.
3. Publicity which places the plaintiff in a false light in the public eye.
4. Appropriation, for the defendant's advantage, of the plaintiff's name or likeness.

Id. at 389.

10. *See generally* Treece, *Commercial Exploitation of Names, Likenesses, and Personal Histories*, 51 TEX. L. REV. 637, 638-41 (1973); Comment, *Transfer of the Right of Publicity: Dracula's Progeny and Privacy's Stepchild*, 22 U.C.L.A. L. REV. 1103, 1103-17 (1975).

theory, a plaintiff had to show not that the appropriation was financially damaging, but that the appropriation caused the plaintiff embarrassment, anguish, and public ridicule. A public figure plaintiff found it nearly impossible to collect damages when his name or likeness was appropriated,[11] since he was seldom embarrassed or subjected to ridicule by truthful, positive publicity. Prominent persons such as actors, athletes, and entertainers seek not to prevent the positive use of their name and likeness, but rather seek the economic value of such use. While courts often recognized a *need* to protect this economic value,[12] they were so entrenched in the "hurt feelings" basis of recovery that they seldom provided actual protection. This judicial trend was, however, reversed in 1953, when the United States Court of Appeals for the Second Circuit coined a celebrity's interest in the publicity value of his name as the "right of publicity" and gave it protection.[13]

---

11. In Cabaniss v. Hipsley, 114 Ga. App. 367, 151 S.E.2d 496 (1966), plaintiff, a showgirl, consented to the taking of a photograph in a "revealing" pose. The photograph, through some mistake, appeared in a local magazine advertisement for a "Playboy Club" where she had never appeared. The court denied recovery of damages, stating that since plaintiff was a "striptease," such a use of her photograph was not embarrassing but rather was occupational publicity. *See also* Hanna Mfg. Co. v. Hillerich & Bradsby Co., 78 F.2d 763 (5th Cir. 1935), where the court stated that "[f]ame is not merchandise," and that "aside from questions of trademark and unfair competition or libel, a public figure might have difficulty in keeping his name and likeness from respectful use by others." *Id.* at 766; Paramount Pictures, Inc. v. Leader Press, Inc., 24 F. Supp. 1004 (W.D. Okla. 1938); Pallas v. Crowley-Milner & Co., 334 Mich. 282, 54 N.W.2d 595 (1952). *See generally* Nimmer, *The Right of Publicity*, 19 LAW & CONTEMP. PROB. 203, 204-10 (1954). Treece, *supra* note 10, states:

> In reality the injury to sensibilities [hurt feelings] concept does not normally meaningfully apply when a person routinely permits advertising uses of his name or picture. Any anger or outrage that he might feel hardly flows from the shock of confronting his likeness in an advertisement. Rather, his injury takes the form of diminished income. The harm resides not in the use of his likeness but in the user's failure to pay.

Id. at 641.

12. In Sharman v. C. Schmidt & Sons, Inc., 216 F. Supp. 401 (E.D. Pa. 1963),plaintiff, a well known basketball player, was denied relief on the theories of defamation and invasion of privacy where his picture was being used by defendant for advertising beer. The court, however, recognized that "[p]ublic figures in the celebrity category have a valuable property right in their name and image." *Id.* at 407. *See also* O'Brien v. Pabst Sales Co., 124 F.2d 167 (5th Cir. 1941).

13. Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866 (2d Cir. 1953). There the court said:

> This right might be called a "right of publicity." For it is common knowledge that many prominent persons (especially actors and ball-players), far from having their feelings bruised through public exposure of their likenesses, would feel sorely deprived if they no longer received money for authorizing advertisements, popularizing their countenances.

Id. at 868. The right of publicity is now commonly recognized in cases of appropriation when the plaintiff is a prominent person. *E.g.*, Motschenbacher v. R.J. Reynolds Tobacco Co., 498 F.2d 821 (9th Cir. 1974); Cepeda v. Swift & Co., 415 F.2d 1205 (8th Cir. 1969); Ettore v. Philco Television Broadcasting Corp., 229 F.2d 481 (3d Cir. 1956); Grant v. Esquire, Inc., 367 F.

The right of publicity is not limited to a person's name and likeness, but extends to a public person's reputation and accomplishments. In *Palmer v. Schonhorn Enterprises, Inc.*,[14] a group of professional golfers sued when their names and biographical profiles were used without their consent in connection with a golf game marketed by the defendants. Plaintiffs admitted that the information used was truthful. Although the information was available public data, the Superior Court of New Jersey granted relief to the plaintiffs, stating that "[i]t is unfair that one should be permitted to commercialize . . . upon another's name, reputation or accomplishments merely because the owner's accomplishments have been highly publicized."[15]

Because the right of publicity has been recognized only recently, the extent of the protection afforded by the right remains unclear.[16] Traditionally, the recovery for appropriation of another's name or likeness from which the right of publicity springs has been limited in three respects: (1) the appropriation must be of a commercial nature; (2) the appropriation must benefit the defendant; and (3) the right to be free from appropriation is subordinate to first amendment rights of freedom of the press.

The first limitation is applicable only when recovery is based on a statute.[17] In jurisdictions recognizing the common law right of freedom from appropriation,[18] protection has not been limited to

---

Supp. 876 (S.D.N.Y. 1973); Uhlaender v. Henricksen, 316 F. Supp. 1277 (D. Minn. 1970); Palmer v. Schonhorn Enterprises, Inc., 96 N.J. Super. 72, 232 A.2d 458 (1967).

14. 96 N.J. Super. 72, 232 A.2d 458 (1967).

15. *Id.* at 462. In Zacchini v. Scripps-Howard Broadcasting Co., 47 Ohio St.2d 224, 351 N.E.2d 454 (1976), the Ohio Supreme Court stated:

> It is this right, a right of exclusive control over the publicity given to his performances, which the plantiff seeks to protect. For a performer, this right is a valuable part of the benefit which may be attained by his talents and efforts, and we think that this right is entitled to legal protection.

Id. at 460. *See also* Ettore v. Philco Television Broadcasting Corp., 229 F.2d 481 (3d Cir. 1956). *But cf.* Gautier v. Pro-Football, Inc., 304 N.Y. 354, 107 N.E.2d 485 (1952) (plaintiff, an animal trainer, denied relief when his performance during half-time of a football game was televised without his consent).

16. Ettore v. Philco Television Broadcasting Corp., 229 F.2d 481, 490 (3d Cir. 1956).

17. Prosser, *supra* note 9, at 401-03; RESTATEMENT (SECOND) OF TORTS § 652C, Comment b (1977). Five states have passed statutes creating a cause of action in cases of appropriation: CAL. CIV. CODE § 3344 (West Supp. 1977); N.Y. CIV. RIGHTS LAW §§ 50-51 (McKinney 1976); OKLA. STAT. ANN. tit. 21, §§ 839.1-839.3 (Supp. 1977); UTAH CODE ANN. § 76-9-405 (Supp. 1977); VA. CODE § 8.01-40 (1977).

18. *See* Pember & Teeter, *Privacy and the Press Since Time, Inc. v. Hill*, 50 WASH. L. REV. 57, 57 n.3 (1974), and Prosser, *supra* note 9, at 386-88. Three states have refused to create judicially the right of privacy, leaving it up to the legislatures to create a statutory cause of action. Brunson v. Ranks Army Store, 161 Neb. 519, 73 N.W.2d 803 (1955); Henry v. Cherry & Webb, 30 R.I. 13, 73 A. 97 (1909); Yoeckel v. Samonig, 272 Wis. 430, 75 N.W.2d 925 (1956).

commercial appropriation but extends to all unauthorized uses of another's name or likeness.[19] Thus, for example, courts have enjoined the unauthorized use of a prominent politician's name by a political party,[20] granted an injunction to prevent a woman from wrongly claiming she was the plaintiff's common law wife,[21] granted a cause of action for cancellation of a birth certificate wrongly naming the plaintiff as the father,[22] and found a cause of action in the unauthorized signing of plaintiff's name to a telegram.[23]

The second limitation, appropriation for the benefit of the defendant, is a natural result of early theories of recovery. Originally, since a person's name and likeness were considered the property of the person, appropriation by another was held to be conversion of property. However, when courts shifted their emphasis from a property rights theory to an injured feelings theory of recovery,[24] celebrity plaintiffs sought, and courts recognized, other theories of recovery, namely unfair competition,[25] and unjust enrichment.[26] Inherent in both theories is an unjust economic benefit derived by the defendant.[27] For example, in *Metropolitan Opera Association v. Wagner-Nichols Recorder Corp.*,[28] a New York Supreme Court, using a theory of unfair competition, enjoined the defendant from further use of the plaintiff's broadcasts without compensating plaintiff. The plaintiff performed radio broadcasts and produced phonographic records through Columbia Records, Inc. Defendant recorded plaintiff's public radio broadcasts and from these produced and sold phonographic records without compensating plaintiff or Columbia Records. The court held that such an appropriation by defendant sufficiently alleged a cause of action for unfair competition. In

19. RESTATEMENT (SECOND) OF TORTS § 652C, Comment b (1977).

20. State *ex rel.* LaFollette v. Hinkle, 131 Wash. 86, 229 P. 317 (1924).

21. Burns v. Stevens, 236 Mich. 443, 210 N.W. 482 (1926).

22. Vanderbilt v. Mitchell, 72 N.J. Eq. 910, 67 A. 97 (1907).

23. Hinish v. Meier & Frank Co., 166 Or. 482, 113 P.2d 438 (1941).

24. See notes 8-11 *supra* and accompanying text.

25. *See, e.g.*, Sinatra v. Goodyear Tire & Rubber Co., 435 F.2d 711 (9th Cir. 1970); Columbia Broadcasting Sys., Inc. v. DeCosta, 377 F.2d 315 (1st Cir. 1967); Lahr v. Adell Chem. Co., 300 F.2d 256 (1st Cir. 1962); Booth v. Colgate-Palmolive Co., 362 F. Supp. 343 (S.D.N.Y. 1973); Loeb v. Turner, 257 S.W.2d 800 (Tex. Civ. App. 1953). *See generally* Nimmer, *supra* note 11, at 210-14.

26. *See, e.g.*, Fairfield v. American Photocopy Equip. Co., 138 Cal. App. 2d 82, 291 P.2d 194 (1955); McQueen v. Wilson, 117 Ga. App. 488, 161 S.E.2d 63 (1968); Cabaniss v. Hipsley, 114 Ga. App. 367, 151 S.W.2d 496 (1966); United States Life Ins. Co. v. Hamilton, 238 S.W.2d 289 (Tex. Civ. App. 1951).

27. "A person who has tortiously used a trade name, trade secret, franchise, profit a' prendre, or other similar interest of another, is under a duty of restitution for the value of the benefit thereby received." RESTAEMENT OF RESTITUTION § 136 (1937).

28. 199 Misc. 786, 101 N.Y.S.2d 483 (Sup. Ct. 1950).

O'Brien v. Pabst Sales Co.,[29] plaintiff, a famous football player, brought action for invasion of privacy when defendant used plaintiff's photograph on a calendar without his consent.[30] While the United States Court of Appeals for the Fifth Circuit denied relief on privacy grounds,[31] the dissent indicated an action could have been based on a theory of quantum meruit because defendant was benefiting from the publication of the photograph without compensating the plaintiff.[32]

Although courts have been increasingly willing to allow recovery for invasion of the right of publicity, they have also been sensitive to the first amendment guarantee of freedom of the press.[33] Traditionally, recovery for appropriation has been limited when the defendant is a member of the news media. A general rule has emerged denying recovery for appropriation when a person's name or likeness is newsworthy.[34] The "newsworthy" immunity has been invoked in cases involving persons who voluntarily thrust themselves into the news and persons who are of celebrity status. In *Man v. Warner Bros., Inc.*,[35] plaintiff, a professional musician, sued for damages when defendant filmed and commercially exploited plaintiff's performance at the Woodstock Festival at Bethel, New York. The federal district court denied plaintiff recovery, calling the Festival a newsworthy event. Relying on *Time, Inc. v. Hill*,[36] the court granted the defendant a limited immunity requiring the plaintiff to show "that the defendant published false material with knowledge of its falsity or in reckless disregard of the truth."[37] This court con-

29. 124 F.2d 167 (5th Cir. 1941).

30. The plaintiff had authorized Texas Christian University to distribute his picture and biographical data to newspapers and magazines generally. Plaintiff did not object to publicity generally, but to publicity associated with selling beer. *Id.* at 168-69.

31. *Id.* at 169.

32. *Id.* at 170. While the dissent referred to the contract theory of quantum meruit, its remarks are equally applicable to the tort theory of imposing liability on a defendant who "reaps where he has not sown." *See* International News Service v. The Associated Press, 248 U.S. 215 (1918).

33. *See generally* Pember & Teeter, *supra* note 18, at 57; Prosser, *supra* note 9, at 410-19; Treece, *supra* note 10, at 654-71.

34. In Leverton v. Curtis Publishing Co., 192 F.2d 974 (3d Cir. 1951), plaintiff, when a child of ten, was struck by a passing car and narrowly escaped death. As a bystander lifted the girl to her feet, a newspaper photographer took her picture, which appeared in a local newspaper the following day. Twenty months later, the photograph appeared in the defendant's magazine, accompanying an article dealing with pedestrian accidents. The court held that the publication of the photograph in the newspaper the day following the accident may have been appropriation, but was privileged by the first amendment, since the photograph was newsworthy. The privilege was not waived by the lapse of twenty months, even though the magazine was sold commercially for profit. *Id.* at 977.

35. 317 F. Supp. 50 (S.D.N.Y. 1970).

36. 385 U.S. 374 (1967).

37. 317 F. Supp. at 51.

sidered the fact that defendants were inspired by a profit motive irrelevant since the Woodstock Festival was a newsworthy event and the motion picture attempted to portray it factually. In *Current Audio, Inc. v. RCA Corp.*,[38] RCA held the exclusive right to the use of Elvis Presley's name, likeness and facsimile signature for advertising purposes, and the exclusive right to manufacture and sell Presley's recordings. Presley gave a news conference shortly before he began a series of concerts. Current Audio recorded the press conference and produced it in the form of a stereo record. A New York Supreme Court refused to grant an injunction prohibiting the marketing of the press conference recording, stating that Current Audio was just as privileged to market this news as was any magazine or newspaper. These two cases illustrate the prior willingness of courts to apply the newsworthy immunity to situations of commercial exploitation when the subject matter involved was legitimate news. The courts had extended the immunity not only to the news media, but also to motion picture and record producers. This immunity afforded publication of newsworthy events was, however, substantially undercut by the United States Supreme Court in *Zacchini.*

II

In *Zacchini v. Scripps-Howard Broadcasting Co.*,[39] the United States Supreme Court, over the dissenting vote of four justices,[40] found Zacchini had a cause of action based upon the right of publicity. The Court disagreed with the Ohio Supreme Court's decision on immunity, stating that no immunity exists for a broadcast of a "performer's entire act without his consent."[41] In overruling the Ohio Supreme Court's grant of immunity to the press, the Court distinguished the cases relied on by the Ohio Supreme Court. The Ohio court had relied most heavily on *Time, Inc. v. Hill.*[42] The United States Supreme Court noted that the *Time* case sought to protect a person's interest in not being publicized in a "false light"[43] and involved the protection of an individual from humiliation and mental suffering. On the other hand, the interest Zacchini sought to protect was the "right of publicity," and involved the protection

38. 71 Misc. 2d 831, 337 N.Y.S.2d 949 (Sup. Ct. 1972).

39. 97 S. Ct. 2849 (1977).

40. Only three Justices dissented on the merits. Justice Stevens would have refused to hear the case on jurisdictional grounds. *Id.* at 2860-61 (Stevens, J., dissenting).

41. *Id.* at 2856-57.

42. 385 U.S. 374 (1967). *See* 97 S. Ct. 2849, 2854-56.

43. 97 S. Ct. at 2855.

of his proprietary interest in his performance.[44] The Court analogized the states' interest in protecting this proprietary right in a performance to the same interest involved in copyright laws.[45] Copyright laws protect the individual from the risk of appropriation inherent in exposing his work to the public by granting to the individual the exclusive right to benefit from his work. The Court had long recognized the importance of the interests protected by these laws.[46] This proprietary interest of a performer in his performance is thus very different from the emotional interest of an individual in his solitude and requires different consitutional analysis.

The Court also noted a difference between the intrusion a "false light" tort made on first amendment freedoms and that made by the "right of publicity."[47] In *Time*, a "false light" case, protection of the interest in solitude required minimizing the publication of damaging information. In *Zacchini,* however, a "right of publicity" case, protection of the interest in publication did not require minimizing dissemination of information, but only limited who could receive the commercial benefit of publication. The *Time* standard[48] for immunizing the media from liability thus did not apply where the plaintiff alleged a "right of publicity."

The Court reasoned that none of its prior decisions had interpreted the first and fourteenth amendments to require a state to grant complete immunity to the press when the media broadcasts the performer's act without consent. At the other extreme, the Court noted that a state may not completely prevent the media from reporting newsworthy facts about a performance.[49] In resolving this conflict between the competing interests of the performer and the media, the Court offered a guideline drawn from the facts of the *Zacchini* case: "Wherever the line in particular situations is to be drawn between media reports that are protected and those that are not, we are quite sure that the First and Fourteenth Amendments do not immunize the media when they broadcast a performer's entire act without his consent."[50] Thus, the result in *Zacchini* allows each state to decide for itself the manner in which the right of

44. *Id.* at 2856.

45. *Id.*

46. Mazer v. Stein, 347 U.S. 201, 219 (1954). *See* 97 S. Ct. at 2857.

47. 97 S. Ct. at 2856.

48. *Time* required that the plaintiff must show that the media had acted in knowing or reckless disregard of the truth before the media would lose its first amendment immunity. *See* Time, Inc. v. Hill, 385 U.S. 374, 390 (1967).

49. 97 S. Ct. at 2856.

50. *See id.* at 2856-57.

publicity is to be protected, but provides the states a guideline for resolving first amendment conflicts with the media.[51]

III

In *Zacchini*, the Court held that a person's right to pursue his commercial interests may sometimes outweigh freedom of the press when those commercial interests involve the person's good will, reputation, and unique creative efforts. Perhaps the Court's policy decision is an appropriate one since purely commercial interests of individuals lie on the periphery of the basic premises behind freedom of the press.[52] However, even accepting the appropriateness of the Court's policy, the decision leaves the limits of its application in confusion. Since the Supreme Court defers the responsibility of balancing the competing interests to the individual states, states must now (1) analyze the significant problems inherent in the Court's guidelines and (2) establish their own workable standard.

A. *Analysis of the Court's Guideline*

1. *Unworkable Guidelines*—The Court recognized that Zacchini's right of publicity conflicted with freedom of the press, but resolved the issue by simple analogy rather than by fully analyzing the facts of the case and the potential impact upon the news media. The majority found the proper analogy to be copyright protection and rejected an analogy to the privacy and defamation cases that protected freedom of the press.[53] This analysis produced an unworkable guideline that provides little guidance for future cases.

The Court's guideline provides that a plaintiff can avoid a first amendment defense by the media if three requirements are met:

(1) The media must "broadcast" the performer's "act;"

(2) The broadcast must be the performer's "entire act;" and

(3) The broadcast must be made without the performer's "consent."

Each of these requirements raises substantial questions.

The first requirement, "broadcasting an act," does not fully define what types of media are subject to the Court's guidelines. By using the term "broadcast," the Court appears to limit the standard to television. Perhaps radio would be included, but the language of the Court does not make this clear. The decision also leaves unclear

51. *Id.* at 2858-59.

52. See notes 64-71 *infra* and accompanying text.

53. See notes 44-48 *supra* and accompanying text.

whether the *Zacchini* guidelines should be applied to a magazine or newspaper that printed, for instance, a comedian's entire nightclub act as news. One may argue that the mere words would not be an "act," since a comedian's act also consists of his timing, mannerisms, and delivery. By the same argument, a radio station's broadcast of a comedian's performance might not be considered an "act" because of the importance of the visual nature of a comedian's performance.[54] Thus, although television, radio, magazines and newspapers are all distinct forms of media, the radio, newspaper, and magazine may circumvent the guideline by failing to broadcast an "act." As a matter of policy, the *Zacchini* rationale should apply to all forms of media when the media makes public the commercial performance of an individual, but the opinion does not indicate whether it would or not.[55]

The second requirement, broadcasting the "entire act," raises three considerations. The first is the problem of what constitutes an entire act. The news media must guess whether such things as an opening fanfare, performance preparations, or curtain calls are within the scope of the entire act.[56] Second, the news media appears to be able to protect itself simply by beginning its news coverage a few seconds after the act begins or stopping the coverage a few seconds before the act concludes. This possibility of subterfuge provides no real protection to a performer. Finally, the news media must be concerned with acts within an act; *i.e.*, whether an individual performance constitutes an "entire act" even though performed within the context of a larger event.[57] For example, the news media must guess whether a single short skit by a performer, such as Marcel Marceau, is an entire act, or whether the entire act consists of all Marceau's short skits performed during a single show.

The third requirement, consent, provides some practical difficulties for the news media.[58] It may take the media so much time

54. The court's analogy to copyright protection illustrates the importance of the visual aspects (timing, mannerisms, and delivery) of a comedian's act. The basic principle of copyright protection is that the expression of an idea, rather than the idea itself, is protected. Timing, mannerisms and delivery provide the substance of "expression." *See generally* Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 U.C.L.A. L. REV. 1180 (1970).

55. For a contrary view, see Note, *Reconciling Red Lion and Tornillo: A Consistent Theory of Media Regulation*, 28 STAN. L. REV. 563 (1976), which urges the need for different standards for broadcast media and the written media because of their unique characteristics.

56. *See* 97 S. Ct. at 2859 & n.1 (Powell, J., dissenting).

57. *Id.* at 2860 n.3.

58. A collateral issue concerns what actions may constitute implied consent or a waiver of consent. Historically, implied consent has been found when the plaintiff seeks or allows publicity, and his name or likeness is used by the defendant for advertising purposes. In

to gain consent that by the time the performance can legally be shown, it will not be shown because it is no longer newsworthy. Such a result would fly in the face of the historically recognized significance of an unrestrained press.[59] The consent requirement of the Court's guideline potentially allows a privately imposed restraint (delaying or refusing to give consent) to outweigh the public benefits of a free press, severely restricting the first amendment guarantee of freedom of the press.

Overall, the *Zacchini* guidelines fail to provide significant guidance and fail to appropriately balance competing interests. In the first two requirements, the Court creates confusion for subsequent cases by failing to define clearly the kinds of media and the kinds of acts to which th guidelines apply. The third requirement—consent—seems to allow a private interest to outweigh a public one.

2. *A Chilling Effect*—The court's decision has the potential of placing a "chilling effect" upon freedom of the press. As the dissent pointed out:

> [T]he decision could lead to a degree of media self-censorship. . . . Hereafter, whenever a television news editor is unsure whether cer-

---

Johnson v. Boeing Airplane Co., 175 Kan. 275, 262 P.2d 808 (1953), plaintiff allowed his photograph to be taken with an airplane under construction. Plaintiff assumed the photograph would be used in the company magazine. The photograph was posted on company bulletin boards without plaintiff's objection, but subsequently appeared in an advertisement in national magazines. The court held that plaintiff was not entitled to recover damages because he had impliedly consented to the use of his picture. *See also* O'Brien v. Pabst Sales Co., 124 F.2d (5th Cir. 1941); Sharman v. C. Schmidt & Sons, Inc., 216 F. Supp. 401 (E.D. Pa. 1963); Pallas v. Crowley-Milner & Co., 334 Mich. 282, 54 N.W.2d 595 (1952); Sidney v. A.S. Beck Shoe Corp., 153 Misc. 166, 274 N.Y.S. 559 (Sup. Ct. 1934).

Another concern is whether consent given to one broadcaster operates in favor of all broadcasters. Cases have held that consent to one advertiser is consent for all advertising purposes. In Cabaniss v. Hipsley, 114 Ga. Ann. 367, 151 S.E.2d 496 (1966), plaintiff, a showgirl, consented to her photographs being taken to advertise her appearance at various nightclubs. The plaintiff was denied recovery of damages when her photograph was used to advertise another showgirl's performance at another club, the court holding that plaintiff consented to all advertising uses of this type, absent a finding of deliberate appropriation. *See also* Wrangell v. C.F. Hathaway Co., 22 App. Div. 2d 649, 253 N.Y.S.2d 41 (1964); Dahl v. Columbia Pictures Corp., 12 Misc. 2d 574, 166 N.Y.S.2d 708 (Sup. Ct. 1957); Sherwood v. McGowan, 3 Misc. 2d 234, 152 N.Y.S.2d 658 (Sup. Ct. 1956); Martin v. Senators, Inc., 418 S.W.2d 660 (Tenn. 1967).

59. In Grosjean v. American Press Co., Inc., 297 U.S. 233 (1933), the court recognized the significance of an untrammelled press:

> The newspapers, magazines and other journals of the country . . . have shed and continue to shed, more light on the public and business affairs of the nation than any other instrumentality of publicity; and since informed public opinion is the most potent of all restraints upon misgovernment, the suppression or abridgement of the publicity afforded by a free press cannot be regarded otherwise than with grave concern.

Id. at 250.

> tain film footage received from a camera crew might be held to portray an "entire act," he may decline coverage—even of clearly newsworthy events, or confine the broadcast to watered-down verbal reporting, perhaps with an occasional still picture. The public is then the loser. This is hardly the kind of news reportage that the First Amendment is meant to foster.[60]

The dissent lists as examples of borderline cases "an event at a local fair, a circus, a sports competition of limited duration (*e.g.*, the winning effort in a ski-jump competition), or a dramatic production made up of short skits."[61]

The Court added to the potential chiling effect by allowing each state to make its own decision as to whether to immunize the news media or give individuals a right of publicity which outweighs freedom of the press.[62] The Court's own guidelines may have little practical effect, since the decision allows fifty varying standards throughout the nation. Since dissemination of news generally involves multistate areas, the news broadcasters—to protect themselves from damage suits—must follow the most restrictive state's standard.

The United States Supreme Court may have made an appropriate value choice in protecting the right of publicity, but the Court's guidelines for avoiding first amendment conflicts are marred with defects which result in an unworkable standard and a chilling effect on the media. As each state develops its own standard, a different approach should be used to ensure protection to the performer, and at the same time, allow the media enough freedom to operate effectively.

B. A Standard for the States

When a state is confronted with the necessity of establishing its own standard, it should first determine under what conditions, if any, individual rights of publicity should outweigh the guarantees of freedom of the press. An historical understanding of the purposes of freedom of the press, as guaranteed by state and federal constitutions,[63] provides a good beginning point. Self-regulation through competing ideas establishes the basic notion of free speech.[64] Citi-

60. 97 S. Ct. at 2859-60 (Powell, J., dissenting).

61. *Id.* at 2860 n.3.

62. *Id.* at 2857-59.

63. The subsequent analysis of the purposes of freedom of the press is based on the United States Supreme Court interpretation of the first amendment. Since the overwhelming majority of states have similar language in their state constitutions, the Supreme Court analysis is applicable to defining the states' interest in guaranteeing freedom of the press.

64. Gertz v. Robert Welch, Inc., 418 U.S. 323, 339-40 (1974).

zens may publish what they want, but the individual reader (and not state law) determines the truth and the worth of the speech. Freedom of the press is thus considered vital to ensure a free and open government.[65] Such a concept, however, may result in some degree of abuse by falsehoods and intrusion into areas not of public concern.[66]

In an attempt to allow free reign to the press, but, at the same time, protect individuals from falsehoods and protect the right to privacy, the United States Supreme Court has established certain limitations on freedom of the press. In *New York Times Co. v. Sullivan,*[67] the Court established a standard for defamation of public figures. Before a public official can collect damages in a defamation suit, he must show that the statements were made with actual malice; that is, the defendant news media had knowledge the statements were false or had published with reckless disregard of the statements' truth.[68] The Court applied a similar standard in an invasion of privacy case in *Time, Inc. v. Hill.*[69] Based on the standard in *Time,* the Ohio Supreme Court adopted a similar standard for the right of publicity in the instant case.[70] However, such a standard may not be appropriate in situations involving the right of publicity because truthfulness or privacy is not an issue when the media broadcasts a performer's act without his consent. Visual and sound recordings minimize the potential for publicizing falsehoods, and performers are generally not interested in protecting their solitude. A media right to publicize the acts of paid performers does little to promote truth or control government and is thus on the periphery of the historical basis for shielding the news media from damage suits.[71] States can, therefore, develop a standard that gives the individual performer greater protection than the traditional defamation or privacy standards provide.

Several alternative approaches of balancing the right of publicity and freedom of the press have been suggested. In *Paulsen v. Personality Posters, Inc.,*[72] a New York Supreme Court held the

65. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 270 (1964).

66. James Madison said that "[s]ome degree of abuse is inseparable from the proper use of everything; and in no instance is this more true than in that of the press." 4 J. ELLIOT, DEBATES ON THE FEDERAL CONSTITUTION 571 (1836).

67. 376 U.S. 254 (1964).

68. *Id.* 279-80.

69. 385 U.S. 374 (1967).

70. 47 Ohio St.2d 244, 251, 351 N.E.2d 454, 461 (1976).

71. The historical basis of freedom of the press would ensure the media's right to comment on the performance, rather than protection against broadcasting a performer's act without his consent.

72. 299 N.Y.S.2d 501 (Sup. Ct. 1968).

right of publicity subordinate to freedom of the press. Paulsen, a television comedian, ran a "sham" campaign for President in the 1968 election. Defendant marketed posters containing the photograph of Paulsen without his consent. Paulsen sought an injunction, claiming defendant was appropriating his comedy routine. The New York Supreme Court disagreed, holding that sham or no sham, Paulsen's bid at election was newsworthy:

> The scope of the subject matter which may be considered of public interest or newsworthy has been defined in most liberal and far-reaching terms. The privilege of enlightening the public is by no means limited to the dissemination of news in the sense of current events, but extends far beyond to include all types of factual, educational and historical data, or even entertainment and amusement, concerning interesting phases of human activity in general.[73]

The New York court then held that the "right of publicity" is a limited one, and has no application where the use of name or likeness is a matter of public interest,[74] and granted defendant complete immunity. Such an absolute immunity standard for any newsworthy entertainment seems clearly inappropriate. The defendant in *Paulsen* was allowed to continue to profit from printing and selling posters of Paulsen when defendant's real intent was to make money from Paulsen's campaign, not merely to report news. Moreover, *Paulsen* demonstrates the weakness of the Supreme Court's guidelines in *Zacchini.* If the *Zacchini* guidelines were applied to the facts of *Paulsen,* Paulsen could not recover because defendant failed to "broadcast" Paulsen's "entire act." A more appropriate standard would have a different starting point. Instead of being concerned with a "broadcast" or an "entire act," courts should first look to the intent of the defendant news media.[75]

If a plaintiff can show that the media had a willful intent to commercialize another's performance under the guise of news, a prima facie case for damages would be established. Evidence establishing a prima facie case could include: (1) a showing that the news media would normally have to pay for the type of performance used;[76] (2) proof that the performance was used outside the context

---

73. *Id.* at 506.

74. *Id.* at 508-09.

75. *See* Zacchini v. Scripps-Howard Broadcasting Co., 97 S. Ct. 2849, 2860 (Powell, J., dissenting); Pavesich v. New England Life Ins. Co., 122 Ga. 190, 50 S.E. 68, 78-79 (1904); Rosemont Enterprises, Inc. v. Urban Sys., Inc., 72 Misc. 2d 788, 790, 340 N.Y.S.2d 144, 147 (Sup. Ct. 1973).

76. As an example of the criterion, a television station could not broadcast the Super Bowl as news. Sporting events, and other similar events, although within the public's interest, must generally be paid for before a medium may broadcast them.

of normal news coverage;[77] (3) evidence that the performance used by the media is outside the scope of a legitimate public interest of the community;[78] and (4) a showing that the media used an exorbitant amount of advertising for the specific performance used, rather than general advertising for the news program as a whole.[79]

The "intent" rationale would protect most performances from appropriation by news media since most performances, because of their nature or length, could not be used within the context of a regular news program or format. Some performances of short duration may still be unprotected because of the difficulty of establishing a willful intent. The Supreme Court decision, however, seems to allow further protection for a performer. The reasoning of the opinion suggests a liability standard based on a showing of an actual economic injury. To limit liability to willful intent, or even reckless conduct, would essentially be the same standard of liability established by the Ohio Supreme Court.[80] Under a strict reading of the United States Supreme Court's opinion, such an intentional standard would be allowable, but not constitutionally required. Therefore, to protect performers unable to show an actual intent by the media to commercialize the performance, states may want to add an additional theory of recovery.

A plaintiff, unable to prove willful intent, might still recover by showing that an actual economic injury resulted from the media's use of the performance.[81] Evidence that could be used to show an actual economic injury would include: (1) decreased attendance at

---

77. The second criterion allows the media to continue to use the type of film footage it normally uses for news. Thus, although the Super Bowl itself could not be shown as news, filmed highlights of the game could be shown, since it is the type of "performance" generally used in the context of a news program.

78. "Legitimate public interest" may be difficult to define. Therefore, this type of evidence should be limied to "news" clearly outside a locality's interest. An example might be the use of a performance from a local fair in South Carolina shown on a local television station in Utah.

79. Admittedly, media often advertise their news programs and occasionally advertise a specific feature on a news program. To use this type of evidence, a plaintiff would need to show the degree of advertising used for the plaintiff's performance is out of proportion to general advertising practices of the media.

80. The Ohio Supreme Court stated:

> The proper standard must necessarily be whether the matters reported were of public interest, and if so, the press will be liable for appropriation of a performer's right of publicity only if its actual intent was not to report the performance, but rather, to appropriate the performance for some private use, or if the actual intent was to injure the performer.

Zacchini v. Scripps-Howard Broadcasting Co., 351 N.E.2d 454, 461 (1976).

81. A standard based on proveable damages or actual economic injury is consistent with the reasoning of legal scholars. *See* Nimmer, *supra* note 11, at 217; Treece, *supra* note 10, at 671-72.

the performances because of the media's use, resulting in a loss of revenue for the performer; (2) cancellation of existing contracts and engagements to perform; and (3) loss of future contracts to perform.

Since this test of liability uses alternative theories, recoverable damages should also be measured by separate theories. First, a plaintiff who shows a prima facie case of media appropriation should recover from the media the cost the media would have had to pay to use the performance commercially.[82] Because of the media's willful intent, punitive damages may also be appropriate. Second, a plaintiff who failed to prove a prima facie case of media appropriation, but who could show that an actual injury resulted from the media's use of a performance, should be compensated for the amount of the proven injury.[83]

A state standard based on intent and actual injury would prevent many of the problems inherent in the *Zacchini* guidelines. In most instances, the media would have more appropriate guidelines to follow, since the standard would be based on the media's intent shown by the general practices of the industry. In the remaining situations, the difficulty of proving actual economic damage should prevent spurious suits. Moreover, such a standard would provide broader protection to the individual performer. The standard would

82. *See* Nimmer, *supra* note 11, at 216. Nimmer accepts unjust enrichment as the correct theory of recovery, stating that "the measure of damages should be computed in terms of the value of the publicity appropriated by defendant rather than, as in privacy, in terms of the injury sustained by the plaintiff." *Id.*

The dissent points out some of the problems of a theory of recovery based on unjust enrichment. 97 S. Ct. at 2859 n.2. If plaintiff's recovery is based on the amount of increased profits, damages will be difficult to prove. Damages recoverable from non-profit organizations, such as public television stations, would become speculative. Further, basing recovery on increased profits or revenues raises fundamental questions about the nature of commercial news media. A commercial defendant operates on the basis of making a profit; but, as the dissent points out, in the short run, the defendant's revenues from advertisements shown during the newscasts are not affected by the content of the news. In the long run, the content of the news program may have an effect on the advertising revenues of the station, but here, as in the case of public television stations, damages recoverable become speculative.

Damages under a theory of unjust enrichment must be based on cost benefits of the defendant rather than increased revenues of the defendant. In other words, plaintiff would receive as damages the amount the media would have had to pay plaintiff to perform commercially or the amount the media would have had to pay other actors or performers to stage a similar performance.

83. Arguably, confusion may arise from combining two different measures of damage. The rationale behind combining the two measures is that in one case, plaintiff is able to show a willful intent, while in the other case, no intent is shown. When intent is shown, damages should be easier to prove. Moreover, the difficulty of showing an actual injury in cases of appropriation without willful intent may prevent spurious lawsuits against news media. To avoid possible confusion, an alternative approach would be to make a showing of actual injury part of the elements of liability under both theories, but allow plaintiff to recover under unjust enrichment if he establishes intent.

apply to all forms of media, rather than television alone, and would protect the performer's economic interest in the media's use of commercially valuable portions of any performance—not merely "entire acts."

WARREN J. LUDLOW

Trans World Airlines, Inc. v. Hardison: A Limitation on the Employer's Duty to Accomodate the Religious Practices of his Employees

Trans World Airlines (TWA) had employed Larry Hardison in a department that operated twenty-four hours per day on each day of the year. Hardison was subject to a seniority system which had been negotiated between TWA and its labor union.[1] This system provided the senior employees with first choice of days off and holidays, and required the junior employees to accept unbid-for openings in the work schedule.

In the fall of 1968, Hardison transferred to a day shift, thereby relinquishing his previously accumulated seniority. Soon thereafter, he joined the Worldwide Church of God, a denomination that requires its members to abstain from all labor not of an emergency nature on specified religious holidays and on the Sabbath, which is observed from sundown Friday to sundown Saturday. When asked to work a Saturday for a vacationing co-worker, Hardison requested an accommodation that would free him from conflict with his Sabbath observance. TWA met with Hardison who proposed alternatives to Saturday work which TWA rejected. TWA authorized the union to arrange a shift trade between Hardison and another employee, but the union refused because the change would have violated the seniority provisions of the collective bargaining agreement.[2] Hardison failed to report for Saturday work and was subsequently discharged.

Hardison brought an action for injunctive relief[3] alleging that his discharge constituted religious discrimination in violation of section 703(a)(1) of Title VII of the Civil Rights Act of 1964.[4] In particular, Hardison relied on the 1967 Equal Employment Opportunity Commission (EEOC) guidelines requiring employers "to make reasonable accommodations to the religious needs of its employees" when this would not result in "undue hardship on the conduct of the employer's business."[5] The district court held that further ef-

1. International Association of Machinists and Aerospace Workers (IAM).

2. Trans World Airlines, Inc. v. Hardison, 432 U.S. 63, 68 (1977). Any shift exchange was subject to a senior worker's right to bump into the opening.

3. Hardison v. Trans World Airlines, Inc., 375 F. Supp. 877 (W.D. Mo. 1974).

4. 42 U.S.C. § 2000e-2(a)(1) (1970). This section proscribes discriminatory employment practices based on an individual's race, color, religion, sex, or national origin.

The Equal Employment Opportunity Commission is an agency created by 42 U.S.C. § 2000e-4 (1970) which has the power "from time to time to issue, amend, or rescind procedural regulations to carry out the provisions of [Subchapter VI-Equal Employment Opportunities]." *Id.* § 2000e-12(a).

5. The guidelines provide in part:

forts on the part of TWA to accommodate Hardison would constitute undue hardship.[6] The United States Court of Appeals for the Eighth Circuit reversed.[7] In *Trans World Airlines, Inc. v. Hardison,*[8] the United States Supreme Court reversed the Eighth Circuit decision, holding that TWA's failure to accommodate further did not constitute a title VII violation.[9]

I. Discrimination and Religion

While the 1964 Civil Rights Act proscribed religious discrimination,[10] it did not define the term. In 1966, the EEOC attempted to provide some certainty in the area of religious discrimination by promulgating guidelines requiring an employer "to accommodate to the reasonable religious needs of employees . . . where such accommodation can be made without *serious inconvenience* to the conduct of business."[11] This guideline expressed the EEOC's disfavor for employer policies that, while neutral on their face, adversely affected employees whose religious beliefs mandated abstention from work on the Sabbath and other specified days.[12] In 1967, the EEOC promulgated more demanding[13] guidelines requiring an employer

---

(b) The Commission believes that the duty not to discriminate on religious grounds, required by section 703(a)(1) of the Civil Rights Act of 1964, includes an obligation on the part of the employer to make reasonable accommodations to the religious needs of employees and prospective employees where such accommodations can be made without undue hardship on the conduct of the employer's business. Such undue hardship, for example, may exist where the employee's needed work cannot be performed by another employee of substantially similar qualifications during the period of absence of the Sabbath observer.

(c) Because of the particularly sensitive nature of discharging or refusing to hire an employee or applicant on account of his religious beliefs, the employer has the burden of proving that an undue hardship renders the required accommodations to the religious needs of the employee unreasonable.

29 C.F.R. §§ 1605.1 (b), (c) (1976).

6. 375 F. Supp. at 889.

7. Hardison v. Trans World Airlines, Inc., 527 F.2d 33 (8th Cir. 1975).

8. 432 U.S. 63 (1977).

9. *Id.*

10. Congress passed title VII primarily in response to pervasive racial discrimination. *See* Edwards & Kaplan, *Religious Discrimination and the Role of Arbitration Under Title VII,* 69 Mich. L. Rev. 599, 599-600 (1971). Religion as a source of potential bias was given only cursory consideration in committee. *See* 110 Cong. Rec. 1528-29 (1964) (remarks of Rep. Rodino).

11. 29 C.F.R. § 1605.1(a)(2), 31 Fed. Reg. 8370 (1966) (emphasis added).

12. B. Schlei & P. Grossman, Employment Discrimination Law 118 (1976).

13. The 1966 guidelines allowed an employer to establish a normal work week and foreseeable overtime requirements and, in the absence of an intent to discriminate, was not required to alter them to accommodate the religious practices of an employee who had accepted a job aware of any potential conflict. 29 C.F.R. § 1605.1(b)(3), 31 Fed. Reg. 8370 (1966). The 1967 guidelines excluded this provision and, in addition, placed upon the em-

"to make reasonable accommodations to the religious needs of employees . . . where such accommodations can be made without *undue hardship* on the conduct of the employer's business."[14]

These guidelines were soon subjected to scrutiny by the courts. In *Dewey v. Reynolds Metal Co.*,[15] an employee, in accordance with his religious beliefs, refused to work on Sundays as required by the compulsory overtime provisions of a collective bargaining agreement. He also felt constrained by his beliefs not to request another employee to replace him on Sunday, a procedure authorized by the company. The Sixth Circuit found that the employer had accommodated Dewey as required by the 1966 EEOC guidelines[16] by making available a replacement system even though he could not, in good conscience, use it. Going further, the court noted that even if the more exacting 1967 guidelines were applied, Reynolds had reasonably accommodated its employee. The court stated that to accommodate Dewey further would, in effect, discriminate against the majority of the employees bound by union contracts.[17] The court also questioned "the authority of EEOC to adopt a regulation interfering with the internal affairs of an employer, absent discrimination."[18] On writ of certiorari, the United States Supreme Court, equally divided, affirmed the Sixth Circuit's decision.[19]

In 1972 Congress added section 701(j) to title VII which incorporated the 1967 EEOC guidelines on religious discrimination. It states: "The term 'religion' includes all aspects of religious observance and practice, as well as belief, unless an employer demonstrates that he is unable to reasonably accommodate to an employee's or prospective employee's religious observance or practice without undue hardship on the conduct of the employer's business."[20]

The legislative history of section 701(j) indicates Congressional dissatisfaction with the judicial treatment of the EEOC guidelines. In presenting the amendment, Senator Randolph referred particu-

---

ployer the burden of proving the undue hardship that would preclude accommodation. *Id.* § 1605.1(c) (1976). See note 5 *supra.*

14. 29 C.F.R. § 1605.1(b) (1976) (emphasis added). See note 5 *supra.*

15. 429 F.2d 324 (6th Cir. 1970), *aff'd mem., by an equally divided court,* 402 U.S. 689 (1971).

16. The 1966 EEOC guidelines were in effect at the time of Dewey's discharge.

17. 429 F.2d at 330-31.

18. *Id.* at 331 n.1.

19. Dewey v. Reynolds Metal Co., 402 U.S. 689 (1971). The guidelines were afforded similar treatment in Riley v. Bendix Corp., 330 F. Supp. 583 (M.D. Fla. 1971), a case with facts similar to *Dewey.*

20. 42 U.S.C. § 2000e(j) (Supp. V 1975).

larly to problems encountered by Sabbatarian employees when their work schedules conflicted with their religious practice.[21] He further expressed disappointment with recent court decisions, specifically referring to the equal division of the Supreme Court in *Dewey*. Section 701(j) was thus intended to "resolve by legislation . . . that which the courts apparently have not resolved."[22]

The circuit court decisions rendered after the 1972 amendment interpreted the amendment broadly and applied it enthusiastically;[23] the Eighth Circuit, in *Hardison v. Trans World Airlines, Inc.*,[24] going as far as any court in narrowing the boundaries of undue hardship. The court implied that the overtime costs and some inconvenience to the employers, which was necessary to accommodate Hardison, did not constitute undue hardship.[25] TWA's defense that the obdurate position maintained by the union in view of its collective bargaining agreement prevented a trade of shifts met a similar fate. The court stated that a seniority provision which forestalled any reasonable accommodation of an employee's religious observance was "prima facie evidence of union and employer culpability under the Act."[26] The opinion also appeared to limit the duty of the

21. 118 CONG. REC. 705 (1972) (remarks of Sen. Randolph).

22. *Id.* at 705-06.

23. The Fifth Circuit responded first in Riley v. Bendix Corp., 464 F.2d 1113 (5th Cir. 1972), by reversing the district court decision, and concluding that the 1972 amendment and its legislative history clearly indicated that the EEOC guidelines expressed Congressional intent. 464 F.2d at 1116. *Accord,* Cooper v. General Dynamics, 533 F.2d 163 (5th Cir. 1976); Draper v. United States Pipe & Foundry Co., 527 F.2d 515 (6th Cir. 1975); Cummins v. Parker Seal Co., 516 F.2d 544 (6th Cir. 1975), *aff'd mem., by an equally divided court,* 429 U.S. 65 (1976), *vacated and remanded on rehearing,* 97 S. Ct. 2965 (1977); Yott v. North American Rockwell Corp., 501 F.2d 398 (9th Cir. 1974). *But see* Reid v. Memphis Publishing Co., 521 F.2d 512 (6th Cir. 1975). The Sixth Circuit had dealt with the issue earlier in Reid v. Memphis Publishing Co., 468 F.2d 346 (6th Cir. 1972), and held that section 701(j) validated the EEOC guidelines. When the case was again granted appeal, a different panel of judges held that section 701(j) could not establish the intent of the 1964 statute not expressed in that statute. 521 F.2d at 520.

24. 527 F.2d 33 (8th Cir. 1975).

25. TWA's complaint that to accommodate Hardison within the framework of the collective bargaining agreement would result in overtime costs drew no sympathy from the unanimous court. While agreeing with the trial court's observation that reasonable accommodation does not require the employer to "make every effort short of going out of business," 375 F. Supp. at 889, the court implied that overtime costs would not be undue hardship to a company the size of TWA and held that the EEOC regulation "does not preclude some cost to the employer any more than it precludes some degree of inconvenience to effect a reasonable accommodation." 527 F.2d at 40.

26. 527 F.2d at 41. The court regarded this statement to be dictum because of its finding that TWA did not request and therefore the union did not refuse to consider a variance from the seniority provision. The court went further and remarked that if "Saturday work inevitably falls to the employees with the lowest seniority . . . [i]t is no answer to [a Sabbatarian], or to the statute itself, that if he compromises his religious beliefs for a time he may develop enough seniority to practice them again." *Id.* at 41-42 n.12.

employee to cooperate in attempts to accommodate him.[27]

II. HARDISON AND THE SUPREME COURT

Though equally divided in prior cases on the duty to accommodate,[28] the United States Supreme Court reversed the Eighth Circuit in *Hardison* by a seven justice majority and concluded that: (1) the duty to accommodate does not require an employer to take action which conflicts with an otherwise valid collective bargaining agreement; (2) TWA's efforts constituted reasonable accommodation as required by the EEOC guidelines; and (3) to deny other employees their contractually conferred shift preference or to incur more than de minimus costs to accommodate Hardison would discriminate against the majority of TWA's employees.

A. The Seniority System

The Supreme Court held that the duty to accommodate did not require TWA to infringe on its collective bargaining agreement with the union. The Court found support for this proposition in section 703(h)[29] of the 1964 Act which approves "bona fide" seniority systems not the result of an intention to discriminate. This sanction of seniority systems follows the trend of recent Supreme Court rulings. In *United Airlines, Inc. v. Evans,*[30] the Court held that section 703(h) protects seniority systems that give present effect to post-Act discrimination not the subject of a timely charge.[31] Relying on *Evans* in part, in *International Brotherhood of Teamsters v. United*

---

27. Hardison, shortly before formal induction into the Worldwide Church of God, had transferred to another TWA building with a separate seniority list. TWA maintained that Hardison's transfer was uncooperative because he had accumulated sufficient seniority in his previous job to be able to avoid Saturday work. The court held that restricting Hardison's right of transfer was not accommodation, but discrimination. Non-cooperation would result when an employee is unresponsive to a reasonable attempt to accommodate. *Id.* at 39.

28. Cummins v. Parker Seal Co., 429 U.S. 65 (1976); Dewey v. Reynolds Metal Co., 402 U.S. 689 (1971).

29. Notwithstanding any other provision of this subchapter, it shall not be an unlawful employment practice for an employer to apply different standards of compensation, or different terms, conditions, or privileges of employment pursuant to a bona fide seniority or merit system, . . . provided that such differences are not the result of an intention to discriminate because of race, color, religion, sex, or national orgin. . . . 42 U.S.C. § 2000e-2(h) (1970).

30. 431 U.S. 553 (1977).

31. In *Evans,* plaintiff was forced to resign for violating United Airlines' rule banning married flight attendants. Plaintiff failed to file a timely charge against United, but United's rule was subsequently found to be an unlawful employment practice under title VII. Plaintiff was later rehired with no seniority rights. She brought suit claiming that the seniority system gave a present effect to United's post-Act discriminatory practice and was thus itself a present violation of section 703(a)(1).

States,[32] the Court went further and held that "703(h) on its face immunizes all bona fide seniority systems, and does not distinguish between the perpetuation of pre- and post-Act discrimination."[33] These opinions affix a Supreme Court stamp of approval on any seniority system not adopted with the intent to discriminate[34] and which has "been maintained free from any illegal purpose."[35]

While the legislative history of section 703(h) shows some intent to allow the operation of seniority systems that continue the effects of pre-Act discrimination,[36] the Court's decision to protect those that perpetuate post-Act discrimination is tenuous. The Court itself noted, in *Franks v. Bowman Transportation Co.*,[37] that in light of the legislative history, section 703(h) is apparently aimed at defining "what is and what is not an illegal discriminatory practice in instances in which the post-Act operation of a seniority system is challenged as perpetuating the effects of discrimination occurring *prior* to the effective date of the Act."[38] The legislative history further reveals that opponents of title VII were concerned that present workers would be, in effect, punished for enjoying the fruits of employer or union practices that, at the time employed, were discriminatory but not unlawful.[39]

---

32. 431 U.S. 324 (1977).

33. *Id.* at 348 n.30. Prior to *Teamsters*, the overwhelming weight of authority was of the view that 703(h) did not immunize seniority systems that perpetuated the effects of prior discrimination. Thus, such seniority systems were not "bona fide" and themselves violated the Act. This view was first adopted in Quarles v. Phillip Morris, Inc., 279 F. Supp. 505 (E.D. Va. 1968). *See, e.g.*, Local 189, United Paperworkers v. United States, 416 F.2d 980 (5th Cir. 1969); United States v. Sheet Metal Workers Local 36, 416 F.2d 123 (8th Cir. 1969).

34. *Id.* at 346 n.28.

35. *Id.* at 356. Retroactive seniority may be awarded in certain cases for a post-Act violation that has been subject to a timely charge. Franks v. Bowman Transp. Co., 424 U.S. 747 (1976).

36. *But see* Cooper & Sobol, *Seniority and Testing Under Fair Employment Laws: A General Approach to Objective Criteria of Hiring and Promotion*, 82 HARV. L. REV. 1598, 1607-14 (1969); Ross, *Reconciling Plant Seniority With Affirmative Action and Anti-Discrimination*, 28 N.Y.U. CONF. ON LAB. 231, 243-49 (1976).

37. 424 U.S. 747 (1976).

38. *Id.* at 761 (emphasis added).

39. Some legislators feared that racial imbalance might be equated with racial discrimiation so as to require the firing of whites in order to hire blacks, 110 CONG. REC. 7091 (1964) (remarks of Sen. Stennis), or that blacks would be awarded superseniority enabling them to make a quantum leap over white workers with established seniority rights. H.R. REP. No. 914, 88th Cong., 1st Sess. 71-72 (1963). Indicative of the particular fears harbored by the Act's opponents is an interpretive memorandum authored by Senators Clark and Case in an apparent attempt to assuage those fears:

> Title VII would have no effect on established seniority rights. . . . Thus, for example, if a business has been discriminating in the past and as a result has an all-white working force, when the title comes into effect the employer's obligation would be simply to fill future vacancies on a non-discriminatory basis. He would not be

The legislators, however, did not intend to sanction seniority systems that constituted a *source* of discrimination. Religious discrimination is defined by section 701(j) as a failure to accommodate.[40] Therefore, a seniority system which prevents the employer from accommodating the religious preferences of its employees, constitutes a source of discrimination. Seniority systems that force employers to discriminate religiously should not be included within the protection of section 703(h), which is at most designed to protect seniority systems that only carry forward the effects of discrimination. The Court's reliance on *Evans* and *Teamsters*[41] to sanction virtually all seniority systems, therefore, extends section 703(h) beyond the scope intended by Congress. If the seniority system in *Hardison* is to stand, it cannot logically do so solely on an accurate interpretation of section 703(h).

B. TWA's Accommodation

The Supreme Court found that TWA's efforts to accommodate Hardison were reasonable. They based their conclusion on the trial court's finding that TWA held several meetings with Hardison, authorized the union to seek a replacement for him, permitted him to observe his religious holidays, and attempted to find him another job.[42] On their face, these efforts pale in comparison to the extent of accommodation required by the various circuits that have ruled on the matter,[43] yet they belie the insignificance of TWA's attention to Hardison's dilemma. The trial court mentioned only three meetings in which TWA participated and one of these was concerned with Hardison's discharge hearing. At the other two, it appears that TWA's representative merely authorized the union to accommodate.[44] The majority of the meetings referred to in the trial court

---

obliged—or indeed, permitted—to fire whites in order to hire Negroes, or to prefer Negroes for future vacancies, or, once Negroes are hired, to give them special seniority rights at the expense of white workers.

110 CONG. REC. 7213 (1964). *See also id.* at 7207 (memorandum prepared by the Department of Justice arriving at the same conclusion).

40. See note 29 *supra.*

41. 432 U.S. at 82, 83 n.13.

42. *Id.* at 77.

43. *See, e.g.,* Cooper v. General Dynamics, 533 F.2d 163 (5th Cir. 1976) (extended duty to accommodate to labor unions and held that title VII protected not only Sabbatarians but those who, on religious grounds, refused to affiliate with or pay support to labor unions); Cummins v. Parker Seal Co., 516 F.2d 398 (6th Cir. 1975), *aff'd mem., by an equally divided court,* 429 U.S. 65 (1976), *vacated and remanded on rehearing,* 97 S. Ct. 2965 (1977) (employer must show that accommodation will result in "chaotic personnel problems"); Draper v. United States Pipe & Foundry Co., 527 F.2d 515 (6th Cir. 1975) (suggested that employer should actually attempt accommodation rather than forecast hypothetical hardships).

44. 375 F. Supp. 884-85.

opinion involved only the union. Also, the trial court went no further than to state that TWA agreed to attempt to find Hardison another job;[45] no mention is made of an actual attempt or of any success. Finally, TWA's agreement to any adjustments in shift assignments that the union might make amounted to nothing more than an authorization. The trial court found that TWA did nothing more to effect a shift change and the union admitted at trial that it made no real effort to change Hardison's shift.[46]

On the other hand, the alternatives suggested by the Eighth Circuit, such as a four-day week for Hardison or the payment of premium overtime pay to replacements were rejected by the Court as involving undue hardship. The Court did not explore further alternatives, such as allowing Hardison to pay the overtime costs incurred by TWA, either through salary reduction or overtime work at regular pay; or allowing him to retransfer to his former job where he had accumulated sufficient seniority to practice his beliefs.

The conclusions arrived at by the Court concerning the scope of accommodation conflict with both the EEOC guidelines and the Act. TWA's actions on Hardison's behalf amounted to little more than authorizing the union to accommodate. The guidelines and the Act require more. The 1967 guidelines state that "undue hardship, for example, may exist where the employee's needed work cannot be performed by another employee of substantially similar qualifications during the period of absence of the Sabbath observer."[47] The facts show that Hardison's job could be performed by each of two hundred other TWA workers.[48] The guidelines, therefore, contemplate the very situation encountered in *Hardison,* and suggest that TWA should have substituted another employee.

Furthermore, the 1972 amendment is clearly a direct response to *Dewey.*[49] Congress apparently felt that the Reynolds Metal's replacement provision was insufficient accommodation, yet TWA did not even take this step.

C. *Rights of the Majority*

The Court also determined that any further accommodation of Hardison would result in discrimination against the other employees. It found that to deny other workers their contractually vested

45. *Id.* at 884, 888.
46. *Id.* at 888.
47. 29 C.F.R. § 1605.1(b) (1976). See note 5 *supra.*
48. 527 F.2d at 34.
49. See text accompanying notes 17-25 *supra.*

shift preference or, indeed, to incur any more than de minimus cost, when no such cost was incurred to give other employees their desired days off, would involve unequal treatment on the basis of religion in a manner not intended by the drafters of title VII.[50]

This argument also runs afoul of the Congressional response to *Dewey*. The Court in *Dewey* had also expressed concern that allocation of shift preference would discriminate against the majority. The *Hardison* Court, however, took a step further and proscribed not only direct preference, wherein Hardison would be allocated Saturday off in the place of a more senior employee, but also indirect preference involving costs incurred toward the accommodation of Hardison that are not incurred for other employees in allowing their shift preferences. Far from not contemplating such unequal treatment, Congress, by enacting section 701(j), clearly required the preferences prohibited by the Court.

III. Preserving the Legislative Intent of Section 701(j)

If read broadly, the Court's interpretation of the 1967 EEOC guidelines can effectively emasculate section 701(j). Therefore, if the legislative intent behind the 1972 amendment is to be given any weight at all, the opinion should be narrowly construed.

Hardison can be viewed as continuing the Supreme Court's recent approval of seniority systems not themselves the result of a purpose to discriminate. The Court specifically examined only three potential accommodations; a four day work-week wherein a replacement from another department would be utilized; the payment of overtime to an off-duty replacement; and a shift swap which would involve a breach of the collective bargaining agreement.[51] Overtime costs and replacements from other departments, neither of which violated the union contract, were apparently undue hardships because the district court found them so.[52] Thus, these limitations could be read as findings of fact, not holdings of law. Furthermore, the Court's statement "that to bear more than *de minimus* cost . . . is undue hardship" is hardly definitional, and invites circularity if all that is not de minimus is undue hardship. The Court's solid backing of seniority provisions in collective bargaining agreements thus remains as the controlling principle.

Evidence in support of a conclusion that seniority rights controlled the outcome in *Hardison* can be found in the Court's prior

50. 432 U.S. at 81.
51. *Id.* at 84.
52. *See id.* at n.15.

decision in *Cummins v. Parker Seal Co.*[53] *Cummins* involved facts quite similar to *Hardison,* with the exception of the seniority provision. In *Cummins,* the Supreme Court, equally divided, affirmed the Sixth Circuit's decision that the employer had not accommodated the employee's religious practices. Since two justices dissented in *Hardison,* the *Cummins* affirmation may indicate that two other justices differentiated *Hardison* on the basis of the collective bargaining issue.[54] The Court recently granted a rehearing in *Cummins* and vacated both its affirmance and the circuit court decision and remanded the case to the Sixth Circuit for proceedings in light of *Hardison.*[55] That court's decision on remand will undoubtedly shed light on the implications of *Hardison.*

IV. Section 701(j) and the Establishment Clause

A narrow reading of *Hardison,* such as the one suggested above, assumes that an employer's duty to make at least a reasonable attempt to accommodate would remain intact. Language in the opinion, however, intimates that the Court may be wary of requiring too much accommodation, either by the union or the employer. Previously, the Court has not gone beyond section 703(h) to support seniority systems claimed to be discriminatory.[56] In *Hardison,* however, the Court took pains to point out that the rights of the majority were at issue in any accommodation involving infringement upon collective bargaining rights.[57]

Furthermore, the Court paid scant attention to the legislative history of both sections 701(j) and 703(h). This new concern for majoritarian rights and the somewhat tortured readings of the statutes in question[58] suggest that a justifiable interpretation of the

53. 429 U.S. 65 (1976).

54. Marshall and Brennan dissented in *Hardison.* The memorandum decision in *Cummins* did not list the votes of the justices.

55. 433 U.S. 903 (1977).

56. *E.g.*, United Airlines, Inc. v. Evans, 431 U.S. 553 (1977); International Bhd. of Teamsters v. United States, 431 U.S. 324 (1977).

57. The Court has not previously expressed such concern over the rights of union members. In Franks v. Bowman Transp. Co., 424 U.S. 747 (1976), while holding that retroactive seniority could lawfully be awarded to victims of unlawful, post-Act discrimination, the Court noted that the seniority rights of other workers are not "indefeasibly vested" and that "[t]his Court has long held that employee expectations arising from a seniority system agreement may be modified by statutes furthering a strong public policy interest." *Id.* at 778. While *Franks* and *Hardison* are readily distinguished, it is unlikely that the Court would maintain that no policy interest is served by the accommodation of an employee's religious practices.

58. Justice Marshall referred to the majority's statutory intepretation in his dissent:

> The Court's interpretation of the statute, by effectively nullifying it, has the singular advantage of making consideration of petitioner's constitutional challenge unnecessary. The Court does not even rationalize its construction on this ground, however, nor

case may well rest upon the first amendment prohibition against laws "respecting an establishment of religion or prohibiting the free exercise thereof."[59] Although not directly addressed in the opinion, the constitutionality of section 701(j)'s accommodation requirement had been questioned prior to *Hardison*,[60] was argued against extensively by TWA in its brief,[61] and may be implicit in the Court's near-nullification of the duty to accommodate.

In *Committee for Public Education and Religious Liberty v. Nyquist*,[62] the Court, set forth a three-pronged test to be used in determining whether a law can withstand constitutional challenge under the first amendment. Under that test, "to pass muster under the Establishment Clause the law in question, first, must reflect a clearly secular legislative purpose, . . . second, must have a primary effect that neither advances nor inhibits religion, . . . and, third, must avoid excessive government entanglement with religion."[63]

The first and third parts of the *Nyquist* test should pose little or no problem to the constitutionality of section 701(j). The Court has seldom struck down a law for lack of secular purpose[64] and it would not appear that governmental confrontations with Sabbatarian plaintiffs would foster the type of entanglement deemed excessive by the Court.[65]

---

could it, since "resort to an alternative construction to avoid deciding a constitutional question is appropriate only when such a course is 'fairly possible' or when the statute provides a 'fair alternative' construction." Swain v. Pressley, 430 U.S. 372, 378 n.11 (1977).

432 U.S. at 90 (Marshall, J., dissenting).

59. U.S. CONST. amend. I.

60. Cummins v. Parker Seal Co., 516 F.2d 544, 554-60 (6th Cir. 1975) (Celebrezze, J., dissenting).

61. Reply Brief for Petitioner at 18-35, Trans World Airlines, Inc. v. Hardison, 432 U.S. 63 (1977).

62. 413 U.S. 756 (1973).

63. *Id.* at 773 (citations omitted). The majority in Cummins v. Parker Seal Co., 516 F.2d 544 (6th Cir. 1975), employed this test and concluded that both the EEOC guidelines and section 701(j) were not inconsistent with the first amendment. 516 F.2d at 551-54. Judge Celebrezze, in dissent, also applied the test and reached the opposite conclusion. *Id.* at 554-59 (Celebrezze, J., dissenting).

64. *See, e.g.*, Committee for Pub. Educ. v. Nyquist, 413 U.S. 756 (1973) (money grants to non-public schools for maintenance and repair supported by a secular concern for health and safety of school children held unconstitutional on other grounds); Walz v. Tax Comm'n, 397 U.S. 664 (1970) (state property tax exemptions for churches reflect valid policy of prohibiting inhibition of beneficial entities in the community); McGowan v. Maryland, 366 U.S. 420 (1961) (Sunday closing laws provided a uniform day of respite from work). *But see* Epperson v. Arkansas, 393 U.S. 97 (1968) (ban against teaching theory of evolution in public school found to stem from sectarian purpose); Abington School Dist. v. Schempp, 374 U.S. 203 (1963) (requirement of Bible reading in public schools not supported by secular purpose).

65. The Court is apprehensive of two types of political entanglement. One involves a

The primary effect aspect of the test is most troublesome. The Court has approved governmental accommodations and benefits, in the past, based on religion, where the result has been an exclusive benefit to the religious practitioner. These accommodations and benefits survived Court scrutiny, however, due to a tension between the establishment clause and the free exercise clause. Where this tension exists, the government must avoid promoting religion on one hand, but must not inhibit it on the other. For example, in *Sherbert v. Verner*,[66] a woman was refused unemployment compensation on the ground that she failed to accept suitable work. Employers would not hire her because of her refusal, on religious grounds, to work on Saturday. The Court found that to allow her the unemployment benefits would not foster the establishment of her religion, but that to deny her the compensation would violate her right of free exercise. And, members of the Amish religion, in *Wisconsin v. Yoder*,[67] were convicted of violating a Wisconsin law compelling attendance at school until age sixteen. The Amish defendants, claiming their religion prohibited high school attendance, refused to send their children to high school. The Court found that, although "the danger that an exception from a general obligation of citizenship on religious grounds may run afoul of the establishment clause"[68] could not be ignored, Wisconsin's law prohibited the free exercise of the Amish religion.[69]

The competition between the free exercise clause and the establishment clause does not exist within the framework of section

---

"comprehensive, discriminating and continuing state surveillance" of the kind that would have been necessry to insure that state aid to church-related schools would be directed towards secular aspects of education. Lemon v. Kurtzman, 403 U.S. 602, 619 (1971). *See also* Tilton v. Richardson, 403 U.S. 672, 688 (1970); Walz v. Tax Comm'n, 397 U.S. 664, 675 (1970). The other concerns the potential for political divisiveness wherein states or communities would divide on issues along religious lines as the Court feared would happen if state aid to parochial schools were allowed. Lemon v. Kurtzman, 403 U.S. 602, 622-23 (1971). The duty to accommodate would not require the government to assume any long-term administrative obligation and the potential for political divisiveness would be greatly diminished by the fact that section 701(j) is federal legislation administered by the EEOC, a federally funded body with broad responsibilities, distinguishing it from the situation in *Lemon* where direct funding to parochial education occurred on the community level. *See* Note, *Is Title VII's Reasonable Accommodations Requirement a Law "Respecting an Establishment of Religion"?* 51 NOTRE DAME LAW. 481, 482-85 (1976).

66. 374 U.S. 398 (1963).

67. 406 U.S. 205 (1972).

68. *Id.* at 220-21.

69. The Court made it clear in both *Sherbert* and *Yoder* that a "compelling state interest" could overcome the mandate of the free exercise clause. *See, e.g.*, Reynolds v. United States, 98 U.S. 145 (1878) (anti-polygamy statute upheld despite its inhibition of the free exercise of the Mormon religion).

701(j). The first amendment limits only federal and state activity;[70] private corporations thus have no constitutional duty to guard free exercise rights, contrary to the obligation imposed on governmental agencies such as the welfare authorities in *Sherbert* or the school authorities in *Yoder*. Therefore, the establishment clause is free from its traditional adversary in this situation, enabling it to check the power of the government to enact legislation advancing religion unrestrained by free exercise considerations.

Other accommodations have been allowed when the benefit to religion was bestowed upon a broad class of which religious organizations were merely members and thus only incidentally aided.[71] In *Everson v. Board of Education,*[72] the Court upheld the plan of a local board of education whereby tax-raised funds were used to reimburse the parents of school children enrolled in public or private non-profit schools for the costs of transporting those children to school on public buses. The Court found that although public money was thus given to the parents of children in parochial schools, the purpose of the plan was "to help parents get their children, regardless of their religion, safely and expeditiously to and from accredited schools."[73] In *Walz v. Tax Commission,*[74] the Court upheld that tax exemptions for property used solely for sectarian purposes, resting its decision on the fact that the exemptions were granted to all religious organizations "within a broad class of property owned by non-profit quasi-public corporations" that the state had determined to be a beneficial influence in community life.[75] Such an analysis also appears implicit in the Court's approval of draft exemptions for conscientious objectors. Section 6(j) of the Military Training and Service Act[76] exempts from military service those who by reason of their "religious training and belief" are conscientiously opposed to participation in war. When section 6(j) was challenged as a violation of the establishment clause, the Court, in *Welsh v. United States,*[77] expanded the term "religious belief" to include all "deeply held moral, ethical or religious beliefs"[78] thus enlarging the potential class of exemptees from one that included only those who objected on theistic grounds.

Section 701(j) does not fall within this category of cases either.

---

70. Cantwell v. Connecticut, 310 U.S. 296, 303 (1939).
71. *See* 62 VA. L. REV. 237, 246-48 (1976).
72. 330 U.S. 1 (1947). *See also* Wolman v. Walter, 433 U.S. 229 (1977).
73. *Id.* at 18.
74. 397 U.S. 664 (1970).
75. *Id.* at 673.
76. 50 U.S.C. app. § 456(j) (1970).
77. 398 U.S. 333 (1968).
78. *Id.* at 344.

Benefits derived from the duty to accommodate are not incidental, but accrue to a class comprising exclusively religious practitioners[79] and in particular to those whose practices conflict with employer work requirements. Such selective aid to religion has not fared well with the Court.[80] In *Nyquist,* a state plan whereby direct financial aid was given for the maintenance and repair of non-public schools attended by the children of low-income parents was challenged on first amendment grounds. Noting that virtually all qualifying schools were supported by the Roman Catholic Church, the Court distinguished prior decisions that allowed incidental benefits to religion and held that the direct benefits to religious schools under the plan violated the establishment clause. Similarly, in *Sloan v. Lemon,*[81] another state plan providing tuition reimbursements to parents of children attending non-public schools was also found violative of the establishment clause. The Court found that the program involved a special economic benefit to a particular class of citizens based on religion, thus distinguishable from previously upheld programs aiding all parents and religion only incidentally. It would thus seem that the duty to accommodate, uncompelled by the free exercise clause, yet preferring only the religious, would fail as having a primary effect that advances religion.

V. CONCLUSION

The Court in *Hardison* left the constitutional door ajar with regard to the scope of required accommodations by employers. A narrow reading of the opinion would still require some duty of accommodation. An attempt by Congress to require more, however, might well force the Court to confront the first amendment issue squarely which, in light of prior decisions coupled with the Court's distaste for the statute, would probably result in the disallowance of any form of preferential treatment for the religious observer in the area of employment.

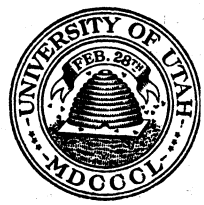
GEORGE PETROW

79. This class is sufficiently broad to cause establishment clause difficulty. The establishment clause not only prohibits aiding one religion, or preference of one religion over another, but also the advancement of all religions. Everson v. Board of Educ., 330 U.S. 1, 15 (1947).

80. *But see* Zorach v. Clausen, 343 U.S. 306 (1952). In *Zorach,* the Court upheld a New York City plan whereby public schools were permitted to release students during school hours in order that they might receive religious instruction away from the school grounds. The law required those students not attending the instructions to remain in the classrooms. This decision appears aberrant, but can be distinguished from the mainstream decisions in that the accommodation was not granted at the expense of taxpayers or employers.

81. 413 U.S. at 825 (1973).

UTAH LAW REVIEW



VOLUME 1977

SALT LAKE CITY, UTAH
1977

INDEX TO VOLUME 1977

ARTICLES — AUTHORS

ARTICLES — TITLES

NOTES

COMMENTS

UTAH LAW

SUBJECT INDEX

HOW EFFECTIVE IS YOUR PERSONNEL RECRUITING PROGRAM IN SELECTING THE RIGHT ATTORNEY FOR YOUR FIRM?

The University of Utah College of Law Placement Office has been established to provide an improved placement program designed to acquaint lawyers, law firms, and other employers of legally trained persons with our high caliber students and alumni, and to enable our students to make informed choices among the wide range of professional opportunities.

The Placement Office encourages on-campus interviews by prospective employers and can assist in making other arrangements, when necessary, to facilitate the employment of its law graduates. The Placement Office maintains resumes which indicate each student's educational background, experience, and areas of interest. Employers wishing to interview and hire law students or recent graduates for summer, part-time, or permanent positions are invited to contact the Placement Office to make the necessary arrangements.

The Placement Office publishes a brochure containing general information about the studentbody, as well as information concerning the various programs and facilities of the College of Law.

Additional information or a copy of the brochure can be obtained by contacting:

THE COLLEGE OF LAW PLACEMENT OFFICE
UNIVERSITY OF UTAH COLLEGE OF LAW
SALT LAKE CITY, UTAH 84112
(801) 581-7767

The University of Utah College of Law, as a member of the Association of American Law Schools, subscribes to and endorses the "Approved Association Policy" of providing equality of opportunity in legal education and employment without discrimination on the grounds of race, color, religion, national origin, or sex. All prospective employers to whom the College of Law furnishes assistance, facilities, or other placement functions are expected to observe these principles of equality of opportunity in their employment practices.